

THIS WEEK

EDITORIALS

AUTHORSHIP Launch of Nature Index to collate citation data **p.164**

WORLD VIEW Survey drug users to find psychiatric medicines **p.165**

EVOLUTION Tracking the crawl of insects through history **p.166**



The burden of depression

Mental health is gaining acceptance as a medical problem, but progress in finding treatments is being hampered by the stigma surrounding people's everyday experiences.

The most common mental disorders are anxiety, depression and substance abuse. The comedian and actor Robin Williams was affected by all three — although he had no drugs or alcohol in his body when he chose to take his own life, as the results of his autopsy showed when released last week.

There was much public discussion in the days and weeks after Williams's death about mental illness, and depression in particular. Details of his suicide and speculation about the pressures he faced went into lurid and undignified detail. Commentators — both professional and amateur — weighed in on the validity (or not) of the image of the tragic clown, and the 'demons' that got the better of him. Then the debate switched to stigma and its reduction (or not). Even to have such a conversation was enough for some. Society was able to give itself a pat on the back for being so open about mental illness.

But how much of that well-meaning discussion focused on the reality that people with depression, and other mental illnesses, face every day? For lack of a better way to put a terrible thing, there seemed to be more focus on the glamorous tragedy of the end, and not so much on the chronic debilitating (and recurring) struggle that came before it. Frankly, it seems that the death of someone with depression is easier to talk about than the life of someone with depression. That needs to change.

Mental illness hits people with a double-whammy. There is the agony of the symptoms themselves: the paralysis of depression, the frustration of psychosis, the panic of anxiety; and then there is the way in which those symptoms interfere with how the person would otherwise like to live his or her life. (Physical illness and injury, too, can bring this secondary effect, of course — see the effect on a sports star kept out of action with a broken leg.)

Better understanding of mental disorders and honest discussion and acceptance of them can help to reduce these secondary effects, but there is further to go here than even the most optimistic advocate would admit. The stigma of living with depression and other mental disorders may have been reduced in recent years, as more people describe their experiences, and wildfire social-media campaigns stamp down on any public statements that could give offence, but the stigma, for example, of being treated for these conditions is harder to shift. Don't agree? Then where are the popular media and political campaigns for fairer and better access to drug and psychological treatments? How come new cancer drugs are heralded as miracle cures, whereas antidepressants and antipsychotics are still whispered about in many countries as part medicine, part magic? Who talks about a struggling diabetic as being 'off her meds'?

The main focus of science and research is to address the symptoms of mental illness. And as this special issue shows, the need to make progress on this goal, for depression in particular, is colossal. Measured

DEPRESSION

A Nature special issue
nature.com/depression

by the years that people spend disabled, depression is the biggest blight on human society — bar none. Research has struggled to lift the fog. And as a result, more than 350 million people still battle with the disorder every day.

Only in the past decade or so have the genomic and brain-imaging tools emerged to enable significant progress on the physical and molecular features of illnesses such as depression. As a result, long-term agendas are shifting, with the US National Institute of

"Neuroscientists and psychiatrists alike must remain open-minded."

Mental Health leading the attempt to partition mental disorder into component symptoms as a way to examine neural mechanisms. Supporters and critics trade ferocious rhetoric about the promise and pitfalls of this approach. And whether it

will capture the most important determinants of mental illness, or enough of them to justify shifting the bulk of research resources, is a legitimate question. There is much criticism that the approach overlooks cultural, social and societal factors. True, but productive study of those — within or outside neuroscience — in ways that translate into clinical benefit has not been cracked with the existing approach either.

Neuroscientists and psychiatrists alike must remain open-minded. Genetic and neuroscience techniques to probe circuits will surely make progress. The fruits of this research will come too late for many people with depression. But, these people still have a crucial part to play. As more people are willing to talk about their experiences of mental illness — and the trend there seems positive — then important insights about the most relevant behaviours and factors will emerge. Discussion of mental illness must continue to grow, and to include the lives of sufferers as well as their deaths. ■

High ambition

The European Space Agency can be proud of its comet mission — whatever happens.

In science-fiction circles, there was excited chatter last week about online trailers for *Ambition*. This slick short film depicts humans in the distant future — somehow imbued with the power to craft their own worlds — discussing how they discovered the key to life.

The film, directed by Oscar nominee Tomek Bagiński and starring Aidan Gillen from the television show *Game of Thrones*, was not what it seemed. It was the work of the European Space Agency (ESA), using

the film as a Trojan Horse to trick an unsuspecting public into engaging with science. The fictional secret to life, the full film revealed, was unearthed with the help of Rosetta — a real ESA mission that this week plans to land a probe on a real comet.

The goal of the mission is more modest than in the film: it hopes to find clues as to whether ancient comets could have delivered Earth's oceans or even the organic molecules that sparked life. Although this is a fascinating puzzle, it is unlikely to be the feature of the mission that people remember. The real drama is the high-stakes attempt to land a probe on the 4-kilometre-diameter comet — 67P/Churyumov–Gerasimenko — which is travelling at more than 60,000 kilometres per hour some 500 million kilometres from Earth. The lander was due to be released, and its fate learned, soon after *Nature* went to press.

The lander, Philae, is intended to provide a ground truth with which to cross-check measurements from the Rosetta orbiter. But it is also designed to go beyond the orbiter's instruments, using an on-board laboratory to analyse samples from 20 centimetres beneath the comet's surface (including materials that would not make it to the dusty tail) and by studying its mechanical properties and interior.

Missions have flown to comets before. Whereas ESA's Giotto and NASA's Stardust provided snapshots, sampling the tail of comets as they raced past at high speed, Rosetta has stalked its prey at close range. Landing the Philae probe was always going to be the riskiest part of the mission. NASA's Deep Impact mission deliberately crash-landed a probe on the surface of a comet to unearth pristine dust for its parent craft to study, but this soft landing is a first.

Unlike asteroids, comets are active bodies that send out streams of water, gas and dust. Rosetta must navigate the harsh environment with sufficient delicacy and precision to launch the lander into a 1-square-kilometre patch on the comet's surface, some 20 kilometres away. Philae must land with a gentle bump, travelling at just 1 metre per second. The slow descent is crucial because the comet's gravitational pull is several hundred thousand times weaker than that on Earth. Even a safe landing is not the final act; the craft must deploy harpoons and ice screws to tie itself to the comet's unforgiving surface.

Experienced ESA flight managers say that this is the most ambitious mission they have ever flown. In terms of complexity, it is more like *Apollo 11*'s landing of humans on the Moon than a Mars adventure. And it is being done with 20-year-old, space-hardy technology, the processing power of which is more like that of a pocket calculator

than anything recognizable on a desktop today. Two decades in the planning (and a decade in transit), the Rosetta mission was developed after NASA cancelled its own Comet Rendezvous Asteroid Flyby mission. The Europeans built on NASA's plans, but instead of a probe, went for the (perhaps more risky) soft-landing option.

Rosetta was always going to be flying into the unknown. All that was known about the comet was its size and orbit around the Sun. Only when it arrived, and discovered 67P to be a bizarre, rugged, rubber-duck-shaped comet rather than the potato the mission scientists were expecting, did the team begin to work out where best to set down the lander.

Rosetta is also a mission of endurance. For a decade, the craft has been chasing its comet through the Solar System, including the cold of space beyond the asteroid belt, for which it was designed to survive temperatures as low as -180°C .

There are lots of ways the mission could (did?) fail, all of which are illustrated in a graphic on page 172. The final manoeuvre ahead of separation, which sets the lander up with the speed and trajectory for its unpowered descent, must go off without a hitch. The craft and lander must separate in perfect synchrony. And then there is luck. Even if Rosetta performs perfectly and mission scientists do everything right — as has been the case so far — if Philae were to land on a cliff, ridge or boulder the probe would topple, and it could be game over.

ESA scientists are not just covering their backs when they say that the mission will be a success whether or not Philae lands safely. Whatever happens to the lander, Rosetta will continue orbiting the comet, the first spacecraft to take a ringside seat as a comet changes in its approach to the Sun. Landing may be the high-drama part of the mission, but around 80% of the science output is expected to come from Rosetta.

Philae's fate might dictate Rosetta's public legacy, but all scientists should celebrate the mission's attempt — whether in homage to its triumph or memorial to its loss — for the engineering and technical expertise it took to execute.

In attempting to capture the public's imagination by stealth, *Ambition* was somewhat out of character for ESA or, indeed, any public research body. But the film — which must have cost a large chunk of the mission's public-relations budget — was probably unnecessary. The feat alone is spectacle enough. ■

ANNOUNCEMENT

Launch of the Nature Index

This week, Nature Publishing Group introduces the Nature Index. The platform is a database of the contributions made by the world's research institutions to articles published in 68 leading scientific journals (see natureindex.com for a freely accessible 12-month data set). The journals were chosen by an independent panel of researchers drawn from across the natural sciences (see natureindex.com/expert-advisers), with further validation from a large online survey of active scientists. The index offers various options for counting and attributing contributions.

The launch is in the spirit of a beta-test — readers can analyse and interpret the data for themselves, and assess the Nature Index's strengths and weaknesses. The methodology behind the index is explained in the accompanying Nature supplement (page S52), as are the caveats that need to be considered when

analysing and interpreting the data.

The journals were selected in 2011. Their outputs have been tracked in subsequent years, and the database will initially be updated monthly. Some journals that might now be strong contenders for inclusion, but were only newly established at that time (including one of the Nature research journals, *Nature Climate Change*), were not selected by the panel. Subject to the feedback we receive, we expect to review the list of journals next year.

We anticipate diverse opinions on many aspects of the Nature Index. Constructive suggestions for improvement are particularly welcome. The supplement includes tables that summarize patterns of research output in the Nature Index for institutions and countries, but these tables are not themselves the Index — rather, they are just some of the possible ways of displaying patterns in a snapshot of data encompassing the most recent complete calendar year.

Used carefully, with proper consideration of the strengths and limits of the underlying data, we believe that the Nature Index can contribute to an understanding of significant scientific outputs at the institutional, country and regional levels. We invite feedback, which should be posted at natureindex.com. ■

TONY BUCKINGHAM/REX



Help luck along to find psychiatric medicines

To increase the chances of stumbling on existing drugs that can double as brain treatments, a systematic search is needed, says **David Nutt**.

Where will new psychiatric medicines come from? That there is a crisis in the drug-discovery pipeline for mental health is well known. Many pharmaceutical companies have given up on this realm, and a research director from one such firm has said that any business working in this field itself needs psychiatric help.

Scientists continue to search for the molecular and neurological clues that could allow informed development of new treatments, but it seems likely that, for now at least, we will have to rely on luck.

Serendipity is not to be sniffed at. The antibiotic minocycline has emerged as a possible schizophrenia treatment after two Japanese people given it for infection saw remarkable improvement in their psychotic symptoms. The blood-clotting drug warfarin may also be useful in schizophrenia: in an audit of clinical work, a US haematology group found that the drug seems to relieve symptoms. There may be many explanations for these observations, and trials are required for confirmation. But if mechanisms can be determined, they could open useful avenues for schizophrenia research.

For depression, clinically relevant observations have been made by neuroscientists exploring brain function. Studies that used ketamine to model psychosis in normal volunteers found that the drug induced rapid improvement in mood. This led to clinical trials that proved its efficacy against depression. Psilocybin, the active psychedelic ingredient of magic mushrooms, has been used to study altered states of consciousness in healthy people, in whom it triggered long-lasting improvements in mood. This, coupled with effects on the subgenual cingulate cortex, an important brain region for depression, has led to a trial on treatment-resistant depression that will start in 2015, funded by the UK Medical Research Council.

Such accidental discoveries are not new: it can be argued that all the commonly used classes of psychiatric drugs have been arrived at through serendipity. The calming effects of the benzodiazepines and lithium were found accidentally in animal studies, and antidepressants and antipsychotics were discovered from unexpected beneficial psychological effects in humans. Early classes of antidepressants were originally developed as anti-tuberculosis and antipsychotic drugs. The mood of patients improved even though the underlying disorder did not. The first antipsychotic, chlorpromazine, showed anti-schizophrenia effects when used as a sedative.

On the basis of such insights, many other agents have been synthesized and tested. Often, these treatments are effective and carry fewer side effects than the original compounds. But they are all 'me too' drugs, based on the same mechanisms of action. So they do not

DEPRESSION

A Nature special issue
nature.com/depression

COMMONLY USED
MEDICINES ARE
CONSIDERED
SAFE,
SO TRIALS FOR
PSYCHIATRIC
BENEFITS COULD BE
FAST-TRACKED.

advance knowledge of the neuroscience of antidepressants or antipsychotics, at least in terms of new molecular targets.

Every day, billions of people take medicines. Some of these drugs may have unexpected effects

on symptoms of brain disorders, but at present there is no way to collect this information. If we are to rely on serendipity to bring us new drugs, how can we increase the chances of success?

Most Western nations already collect information on adverse effects. We should establish parallel schemes to report psychiatric benefits. Trials could be especially useful in countries such as Denmark and Sweden, which have existing databases for medicine use and for mental health. Commonly used medicines are considered safe, so trials for psychiatric benefits could be fast-tracked.

Patient-led approaches are also possible. Online forums allow people to discuss their experiences of antipsychotic and antidepressant medications. Such sites may have millions of visitors per month and would be ideal hosts for large-scale public questionnaires, or web pages dedicated to reporting positive influences of prescription and over-the-counter medicines.

A rather different (and more radical) approach is to tap into the recreational use of non-prescription drugs. Millions of people take these compounds, and not all of the drugs are illegal. Indeed, certain 'legal highs' have chemical similarities with therapeutic drugs and could lift mood or reduce anxiety. Users are often keen to share their experiences, at least through online questionnaires.

Apart from the obvious political barrier to such research (tacit endorsement of illegal drug use), there is a big scientific stumbling block. Illegal recreational drugs are not subject to trading-standards regulations and may not contain what the name suggests. However, legal highs are usually a pure form of what they claim to be.

Insights from public use of drugs and medicines would not be from controlled trials, so formal publication in journals could be difficult. But dedicated databases to showcase findings would be cheap to establish. Perhaps companies that developed treatments on the basis of these discoveries could make a contribution to the costs. Maybe the registers could attract a small part of the intellectual property from any such discoveries. Companies with drugs currently in patent would certainly have a significant financial interest in the compounds being evaluated for new psychiatric indications.

The investment is worth it. The drugs that we need to ease the burden of mental illness could be out there — we just need to look. ■

David Nutt is a psychiatrist at Imperial College London, UK.
e-mail: d.nutt@imperial.ac.uk

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

CHEMISTRY

Controlling the coffee-ring effect

A drop of coffee usually leaves a ring-shaped stain when it dries on a surface, but researchers have now used light to control the shape of the deposit left behind.

In the coffee-ring effect, particles flow to a droplet's edge as the liquid evaporates, forming the ring. To change this, Damien Baigl and Manos Anyfantakis from the Ecole Normale Supérieure in Paris added surfactants to polystyrene nanoparticles suspended in water droplets. The polarity of surfactant molecules changed when exposed to ultraviolet light, sticking the nanoparticles to the top of the droplet. This stopped the particles moving to the edge as liquid evaporated, so that they formed a disc instead of a ring.

The technique could prove useful for controlling deposits when droplets that contain particles dry, such as in ink-jet printing.

Angew. Chem. Int. Ed. <http://doi.org/f2vh3d> (2014)

EVOLUTION

Insect natural history traced

Insects emerged some 480 million years ago, eventually becoming the most diverse animal group.

Details about the early evolution of the more than 900,000 known insect species are still contentious. Bernhard Misof at the Alexander Koenig Zoological Research Museum in Bonn, Germany, and his colleagues determined the

evolutionary relationships of all 30 orders of insects by analysing fossil and molecular data. The first insects evolved in marine or coastal habitats and took to land at around the same time as plants appeared, some 440 million years ago. Flying insects appeared around 40 million years later, followed by a boom in insect diversity.

All the major lineages of insects that are alive today were in existence by about 375 million years ago, giving the animals plenty of time to diversify even more (**pictured**). *Science* 346, 763–767 (2014)

ATMOSPHERIC SCIENCE

Volcanic cooling underestimated

Since 2000, atmospheric particles from volcanic eruptions have cooled the Earth more than scientists had suspected.

Aerosols from volcanoes cool the planet by reflecting sunlight, and are typically tracked by satellites that monitor altitudes of 15 kilometres and higher. A team led by David Ridley, of the Massachusetts Institute of Technology in Cambridge, used laser and aerosol

measurements from the ground, as well as from balloons, to estimate volcanic particles occurring at altitudes of roughly 8–15 kilometres.

They found a high enough level of particles to account for a global cooling of 0.05–0.12 °C since 2000. This could be contributing to the observed pause in global warming since then, the authors say.

Geophys. Res. Lett. <http://doi.org/wxk> (2014)

MICROBIOLOGY

Cell decoys capture toxins

Specially designed lipid sacs bind to deadly bacterial toxins in infected mice, helping the animals to clear the pathogen.

Certain bacteria release toxins that kill host cells by binding to the cell membrane. Eduard Babychuk at the University of Bern, Switzerland, and his team created liposomes — artificial spheres made of cholesterol and other lipids found in cell

membranes — to act as decoys and bind the toxins.

In a lab dish, the liposomes captured various bacterial toxins, protecting human cells. In mice infected with *Streptococcus pneumoniae* and *Staphylococcus aureus*, the liposomes stopped the animals getting blood poisoning, whereas untreated animals died in under two days.

Such liposomes, already in use as drug carriers, could help to treat certain bacterial infections, the team says.

Nature Biotechnol. <http://doi.org/wxm> (2014)



ULET IFANSATI/GETTY



OLIVER NIEHUIS, ZFMK/BONN

CLIMATE CHANGE

Tropical storms not intensifying

The potential intensity of tropical cyclones has not increased in recent decades, contrary to expectations in a warming world.

James Kossin of the US National Climatic Data Center in North Carolina assessed past storm intensity by analysing 30 years of cloud-top temperature data from satellite imagery of tropical cyclones. This removed uncertainties found in other atmospheric analysis techniques. He did not find an increase worldwide in the average potential intensity of storms, even though a warmer climate is thought to support stronger cyclones.

The lack of a trend is largely due to the poleward migration of tropical cyclones into climates that are less able to fuel intense storms than the tropics, Kossin says.

Bull. Am. Meteorol. Soc.
<http://doi.org/ww3> (2014)

ASTRONOMY

Amateurs spot unusual exoplanet

Citizen scientists have found an exoplanet that was overlooked in previous analyses of data from NASA's Kepler satellite.

The planet, about 700 parsecs away, is roughly four times as massive as Earth, according to Joseph Schmitt of Yale University in New Haven, Connecticut, and his colleagues — including three users of the Planet Hunters website, which crowdsources analysis of Kepler data. The planet, named PH3 c, also has a low density, suggesting a substantial atmosphere of hydrogen and helium gas.

The volunteers spotted dips in the brightness of the planet's host star as the planet moved in front of it every 66 days. The gravitational pull of other

planets caused the timing of these transits to vary by 10 hours, allowing the planet to elude Kepler's automated search. The planet is the third confirmed find by the Planet Hunters community.

Astrophys. J. 795, 167 (2014)

NANOTECHNOLOGY

Nanocoils make batteries stretchy

Coils of carbon nanotubes can be used to make lightweight, stretchable batteries.

Huisheng Peng and his team at Fudan University in Shanghai, China, twisted together several nanotube fibres and found that they had high electrical conductivity, even when stretched up to three times their initial length.

The team incorporated lithium-based nanoparticles into the springy fibres to make a lithium-ion battery, which also performed stably after being stretched. The device is smaller and lighter than batteries that are built on elastic substrates, and can store more charge than such stretchy batteries of similar weight.

The device could be used in 'smart' clothing and other applications, the authors say. *Angew. Chem. Int. Ed.* <http://doi.org/f2vxwm> (2014)

COGNITIVE NEUROSCIENCE

Ghostly feelings from brain mix-up

Sensations of a non-existent person nearby might arise from the brain's failure to integrate different body signals.

Olaf Blanke of the Swiss Federal Institute of Technology in Lausanne and his colleagues found that people who reported these feelings often had injuries in the frontoparietal cortex, which integrates different types of sensory and motor information.

The team recreated this effect in healthy volunteers who

SOCIAL SELECTION

Popular articles
on social media

The benefits of being a big name

Scientists develop reputations that often work to their advantage. A study suggests that the presence of a well-known scientist on a list of authors can drive citations of the paper, regardless of the merits of the research — especially soon after its publication. The report rapidly started an online discussion. "How scientists too can be famous for being famous," tweeted Ed Rybicki, a virologist at the University of Cape Town in South Africa. Naupaka Zimmerman, a microbial ecologist at the University of Arizona in Tucson, took to Twitter to ask: "Do we cite papers b/c of what they say, or b/c of who wrote them?"

Proc. Natl Acad. Sci. USA 111, 15316–15321 (2014)



Based on data from altmetric.com.
Altmetric is supported by Macmillan
Science and Education, which owns
Nature Publishing Group.

➔ **NATURE.COM**
For more on
popular papers:
go.nature.com/v8rn7d

were asked to move a robot in front of them using their finger. A second robot behind them mimicked the movements by touching their back. The participants were blindfolded and wore headphones so that they could not see or hear the robots moving.

When the rear robot moved immediately, most participants felt that they were touching themselves on the back with their own finger, even though their arms were stretched forward. But when there was a half-second delay, they felt that the touch was coming from someone, or something, else. The illusions were caused by a mismatch between the expected and actual sensory information, the authors say. *Curr. Biol.* <http://doi.org/wzb> (2014)

AGRICULTURE

Pasture plants are also weeds

Land used for pasture is at risk of becoming a breeding ground for weeds that invade natural areas.

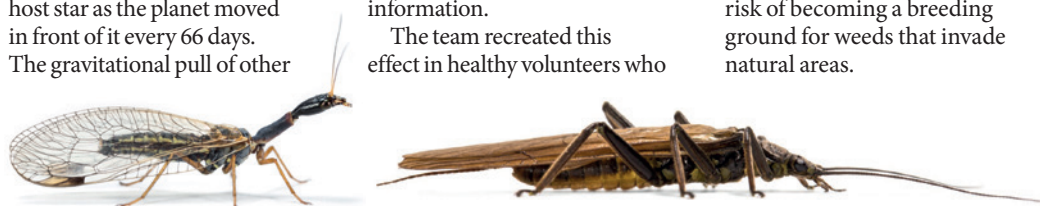
Roughly 30% of land worldwide is devoted to growing forage plants for livestock, and as demand for meat rises, farmers are moving towards more-productive and harderier plants. Don Driscoll at the Australian National University in Canberra and his colleagues looked at data on pasture species promoted by 17 agribusinesses and government agriculture agencies on six continents, and found that 91% of the plants were classified as potentially invasive weeds — often in the same country in which they were developed and marketed. Only one of the 17 institutions had a formal process for identifying possible weeds.

The researchers suggest solutions for this problem, including making the organization that promotes the pasture species financially liable for controlling it if it becomes invasive.

Proc. Natl Acad. Sci. USA <http://doi.org/wxq> (2014)

For a video on the research, see
go.nature.com/gqxoed

➔ **NATURE.COM**
For the latest research published by
Nature visit:
www.nature.com/latestresearch



POLICY

Ebola support

On 5 November, US President Barack Obama requested that Congress approve US\$6.18 billion in emergency funds this fiscal year to fight the Ebola outbreak in West Africa. On the same day, China's state-run media outlet Xinhua reported that the country plans to send about 1,000 medical workers and experts to West Africa in the coming months to help in the fight against Ebola. The Chinese government had previously announced that it would send an elite military medical unit to assist in Liberia.

US election results

Republicans took control of the US Senate in the midterm federal election held on 5 November, winning 52 of 100 seats (with two races yet to be decided). The party maintained its hold on the House of Representatives, taking 244 of 435 seats (with seven races still contested). The Republicans' big win is expected to give key leadership positions to lawmakers who are sceptical of climate-change science, and in favour of smaller federal budgets, in the 114th Congress, which begins on 3 January 2015 and runs to 2016.

Nuclear restart

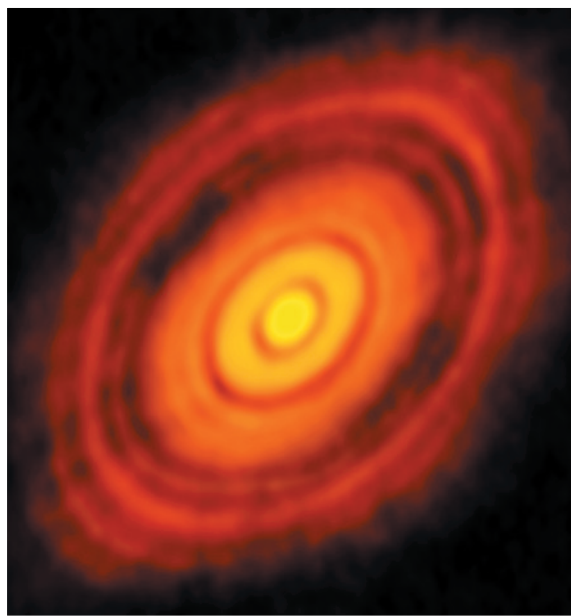
On 7 November, Japan moved closer to resuming nuclear-power generation when a local government approved the restart of two reactors on the southern island of Kyushu. Japan's 48 commercial nuclear reactors have been idle since the triple meltdown in March 2011 at the Fukushima Daiichi nuclear facility. In September, the Nuclear Regulation Authority approved the technical redesign of reactors

at two units of the Sendai Nuclear Power Plant in Kagoshima, Kyushu. Pending the regulator's final approval, the two 890-megawatt pressurized-water reactors could be restarted early next year.

Transgenic potatoes

The US Department of Agriculture decided on 10 November that it did not need to regulate a transgenic potato, after determining that it poses no threat to other

crops. Engineered by Simplot of Boise, Idaho, the potato contains transgenes that silence natural potato genes related to the production of acrylamide — suspected to cause cancer — when the vegetable is cooked. The potato, which also resists bruising, is part of a new generation of transgenic crops aimed at benefiting consumers as well as farmers. Nevertheless, it has met with opposition from some consumer groups.



View of planetary births

This 6 November image may be the best ever taken of a protoplanetary disk, the dust and gas that swirl around a young star and seed new planets. The Sun-like star HL Tau, 138 parsecs (450 light years) from Earth, sits in the centre of an orange-coloured disk of gas and particles that have started to coalesce into clumps. These emerging planets accumulate material as they orbit the star, eventually clearing out the lanes in which they travel, which are seen as black rings. The image comes from the Atacama Large Millimeter/submillimeter Array in Chile, whose 66 radio antennas are being ramped up to image celestial objects in unprecedented detail. Since September, the antennas have been positioned so that their data can be combined to produce the sharpest images yet.

EVENTS

L'Aquila appeal

Six seismologists accused of misleading the public about the risk of an earthquake in Italy were cleared of manslaughter on 10 November. An appeals court overturned their six-year prison sentences and reduced to two years the sentence for a government official who had been convicted with them. The magnitude-6.3 earthquake struck the town of L'Aquila on 6 April 2009, killing more than 300 people. See page 171 for more.

RESEARCH

HIV baby test

The US National Institutes of Health is sponsoring a clinical trial to test whether babies who become infected with HIV in the womb can benefit from antiretroviral therapy given within hours of birth. The phase I/II proof-of-concept study announced last week aims to build on findings from the 'Mississippi baby' case (see *Nature* <http://doi.org/w2n>; 2014), and will test whether early treatment can send the virus into remission. The trial will enrol up to 472 infants and their HIV-infected mothers in at least 9 countries in Africa and North and South America.

FACILITIES

Telescope reprieves

Two astronomical observing facilities won new leases of life last week. On 4 November, the University of California announced that it had reversed a 2013 decision to phase out funding for the Lick Observatory on Mount Hamilton, California (see *Nature* <http://doi.org/w2m>; 2014). Days earlier, the University of Hawaii took ownership of the United

ALMA (ESO/NAOJ/NRAO)

Kingdom Infrared Telescope in Hawaii. In 2012, the UK Science and Technology Facilities Council had announced that it would close the telescope and concentrate funds on a much larger planned telescope in Chile.

Microbiome centre

The Massachusetts Institute of Technology (MIT) in Cambridge is to host a centre for studying the role of the microorganisms found in the human body — known as the microbiome — in health, and for developing treatments for related illnesses. MIT and Massachusetts General Hospital in Boston announced on 6 November that the Center for Microbiome Informatics and Therapeutics will be supported by a US\$25-million fund for its first five years. The centre will initially focus on inflammatory bowel disease, but organizers hope to eventually broaden the scope to diseases such as multiple sclerosis, arthritis and autism.

PEOPLE

CERN chief

Italian physicist Fabiola Gianotti will be the next director-general of CERN, the European particle-physics laboratory announced on 4 November. In July 2012, Gianotti (**pictured**) first revealed evidence for the



existence of the Higgs boson to the world, as spokeswoman for ATLAS, one of two experiments at CERN's Large Hadron Collider that discovered the particle. Gianotti, who will be the first woman to lead the laboratory near Geneva, Switzerland, will take charge on 1 January 2016. See *Nature* <http://doi.org/w2p> (2014) for more.

PubMed leader

Don Lindberg, who has directed the US National Library of Medicine for 30 years, announced on 6 November that he will retire in March 2015. During his tenure, Lindberg created the National Center for Biomedical Information, which hosts such tools as the biomedical-journal database PubMed and the genetic-sequence database GenBank. He was also instrumental

in creating some of the first biomedical computing projects for the White House and other agencies.

BUSINESS

Rocket rethink

Orbital Sciences of Dulles, Virginia, plans to fly cargo to the International Space Station next year using an as-yet unspecified rocket, rather than the Antares design that exploded last month (see *Nature* <http://doi.org/ws5>; 2014). The company announced on 5 November that the accident was probably caused by a turbopump failure in one of its two AJ26 rocket engines, built in the Soviet era and since refurbished. Orbital maintained that it would fulfil its NASA contract to deliver supplies and scientific equipment to the space station by the end of 2016.

COMING UP

15–19 NOVEMBER

Washington DC hosts the annual meeting of the Society for Neuroscience. Programme highlights include research on neurodegeneration, addiction and sleep. go.nature.com/dmfqpf

Animal trouble

Antibody-maker Santa Cruz Biotechnology has fallen foul of the US Department of Agriculture (USDA) for a second time. On 4 November, the agency filed a formal complaint against Santa Cruz of Dallas, Texas, for alleged violations of the Animal Welfare Act. The allegations include failure to provide adequate nutrition, veterinary care and safe, clean enclosures at its animal facilities. In October 2012, USDA inspectors discovered that the company had been keeping an undisclosed herd of research goats, including some ill animals (see *Nature* <http://doi.org/w29>; 2013).

AWARDS

Millionaire prizes

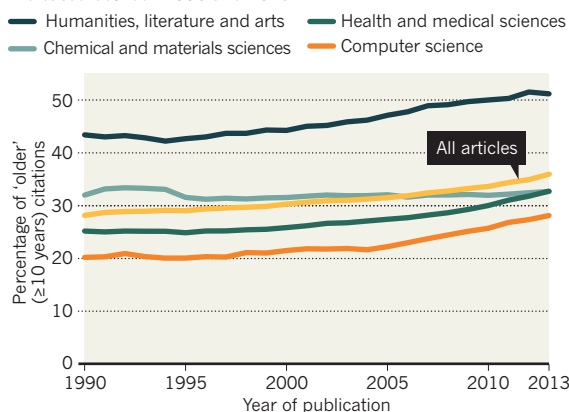
Two teams of astrophysicists shared the US\$3-million Breakthrough Prize in Fundamental Physics, announced on 9 November, for their discovery that the expansion of the Universe is accelerating. The six winners of the life-sciences award each got \$3 million; they included neurosurgeon Alim Louis Benabid, a pioneer of deep-brain stimulation, and biologists who developed the gene-editing technique CRISPR. The awards are sponsored by philanthropists including Google co-founder Sergey Brin. See go.nature.com/dmdcmg for more.

TREND WATCH

An analysis by Google reports that papers are increasingly citing literature more than a decade old (A. Verstak *et al.* Preprint at <http://arxiv.org/abs/1411.0275v1>; 2014), consistent with earlier findings (V. Larivière *et al.* *J. Am. Soc. Inf. Sci. Technol.* **59**, 288–296; 2008). The trend shows that scholars now have easier access to the older literature, the Google researchers say, as print articles have been digitized and the Internet makes it easier to find and cite papers.

GROWING IMPACT OF OLDER ARTICLES

The proportion of citations made to papers at least 10 years old increased between 1990 and 2013.



NEWS IN FOCUS

SPACE What makes landing on a comet so hard **p.172**

MEDICAL ETHICS Rules on clinical trials stir debate **p.174**

NEUROSCIENCE Evidence mounts for gut-brain link **p.178**



DEPRESSION The condition that causes a staggering burden of disability **p.179**

FILIPPO MONTEFORTE/AFP/GETTY



More than 300 people perished in the earthquake that struck the medieval centre of L'Aquila.

RISK ASSESSMENT

Scientists cleared of quake deaths

Italian appeals court says six seismologists were not guilty of manslaughter following the 2009 L'Aquila disaster.

BY ALISON ABBOTT & NICOLA NOSENGO

Six seismologists accused of misleading the public about the risk of an earthquake in Italy were cleared of manslaughter on 10 November. An appeals court overturned their six-year prison sentences and reduced to two years the sentence for a government official who had been convicted with them.

The magnitude-6.3 earthquake struck the historic town of L'Aquila in the early hours of 6 April 2009, killing more than 300 people.

The finding by a three-judge appeals court prompted many L'Aquila citizens to react with rage, shouting "shame" and saying that the Italian state had just acquitted itself, local media

reported. But it comes as a relief to scientists around the world who had been following the unprecedented case with alarm.

"We don't want to have to be worried about the possibility of being prosecuted if we give advice on earthquakes," says seismologist Ian Main of the University of Edinburgh, UK. "That would discourage giving honest opinion."

The defendants themselves have mixed feelings. Giulio Selvaggi, former director of the National Earthquake Centre in Rome, says that although he is happy to be acquitted, "there is nothing to celebrate — because the pain of the people of L'Aquila remains".

In the months before the major earthquake struck, the region around L'Aquila had been

subject to frequent, mostly low-magnitude tremors known as seismic swarms. Residents were alarmed by a local amateur earthquake predictor's claims that he had evidence of an impending quake, although geologists dismissed his methods as unsound.

A commission of experts met on 31 March 2009 to advise the government. According to the prosecution, a press conference after that meeting — attended by the acting president of the commission, volcanologist Franco Barberi of the University of Rome 'Roma Tre', and by Bernardo De Bernardinis, then deputy director of the Italian Civil Protection Department — conveyed a reassuring message that a major earthquake was not on the cards. Moreover, in a television interview recorded shortly before the meeting but aired after it, De Bernardinis said that "the scientific community tells me there is no danger because there is an ongoing discharge of energy" during the seismic swarms.

As a consequence, according to the prosecution, when the earthquake struck on 6 April, 29 people chose to stay indoors and died as their homes collapsed. All members of the expert commission were found guilty of manslaughter in October 2012, after a 13-month trial that transfixed the international scientific community.

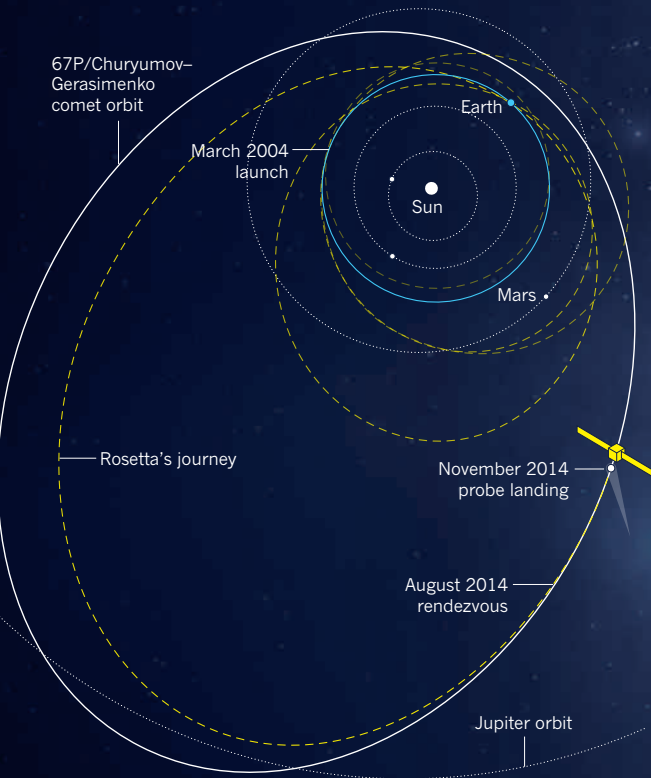
In addition to De Bernardinis, Selvaggi and Barberi, the other defendants were Enzo Boschi, former president of the National Institute of Geophysics and Volcanology in Rome; Claudio Eva, an Earth physicist at the University of Genoa; Mauro Dolce, head of the seismic-risk office of the Civil Protection Department in Rome; and Gian Michele Calvi, director of the European Centre for Training and Research in Earthquake Engineering in Pavia.

Over the course of six hearings before the appellate court in L'Aquila, the defence argued that there was no proof of a causal link between the meeting and the behaviour of the people of L'Aquila citizens. The lawyers also argued that the scientists could not be held accountable for De Bernardinis's reassuring statements, and that their scientific opinions were ultimately correct.

De Bernardinis was acquitted of the manslaughter charges in 16 cases, but not for the other 13. The judges can take up to three months to publish the reasoning behind their verdict. Lawyers for the families of the deceased have announced that they will challenge the ruling in the Supreme Court of Cassation in Rome, which could call for a retrial. ■

LANDING ON A COMET

A STEP-BY-STEP GUIDE TO A PERILOUS DESCENT



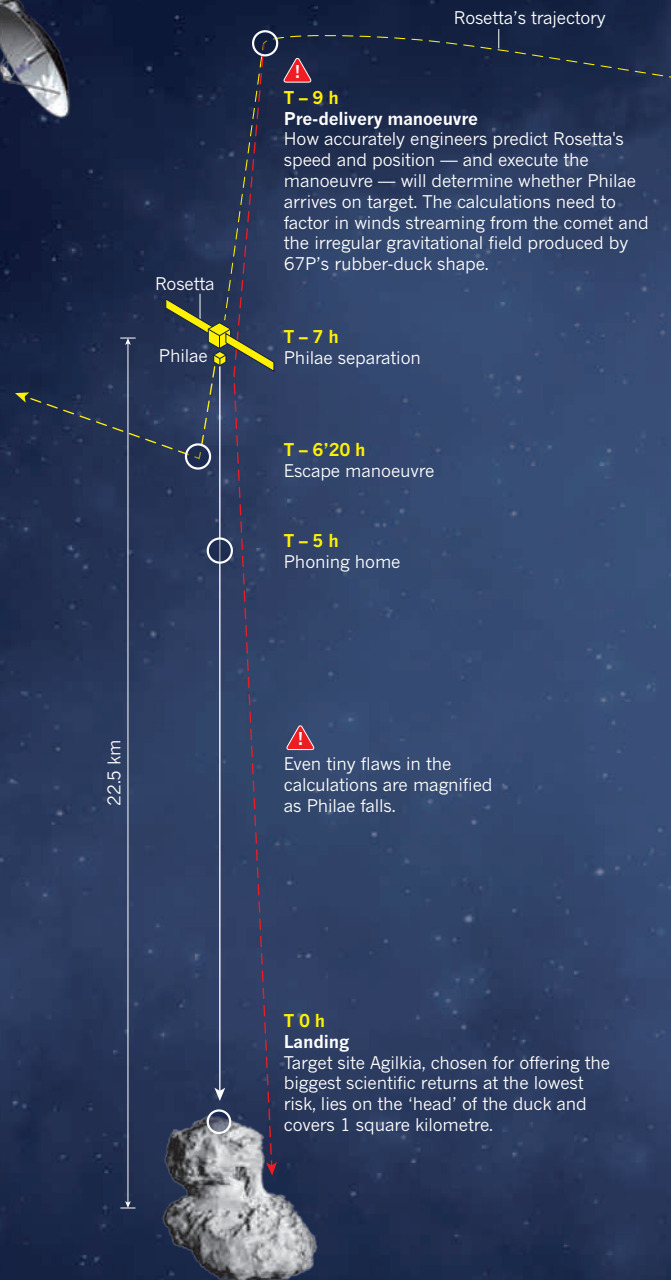
Never before has a space mission put a lander on a comet. But as *Nature* went to press, the European Space Agency (ESA) was poised to change that. Its Rosetta craft had been orbiting comet 67P/Churyumov-Gerasimenko since August and was set to release the washing-machine-sized lander, Philae, on 12 November. This would set in motion a nail-biting seven-hour fall that would deliver Philae to a landing site called Agilkia on the comet's surface. Philae is programmed to beam data and images back to Earth to help scientists to understand comets, including whether these conglomerations of ice, rock and dust supplied our planet with water and other building blocks of life when they smashed into it billions of years ago. But even if the landing fails, it will go down as one of the most ambitious feats attempted in space. Our guide illustrates why.

NATURE.COM
Visit www.nature.com/philae to find out what happened

BYE-BYE MISSION CONTROL

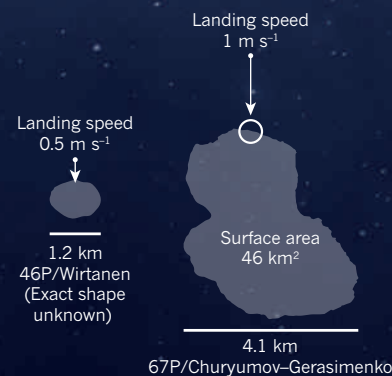
T-9h

Rosetta is steered by infrequent thruster burns, directed from mission control in Darmstadt, Germany. The last burn before landing comes when Rosetta turns towards the comet.



COMET SWITCHEROO

ESA originally planned for Rosetta to visit a smaller comet, 46P/Wirtanen, but a postponed launch forced the agency to choose a replacement. The bigger comet increases the distance at which separation must occur, decreasing Philae's landing precision. It also has a stronger gravitational pull, making the landing heavier.



SYNCHRONIZED SEPARATION

T-7h

Following a cue from a control centre in Cologne, Germany, Philae ejects from the mother ship and begins to go it alone. Forty minutes later, Rosetta pulls back and heads for a more distant orbit.

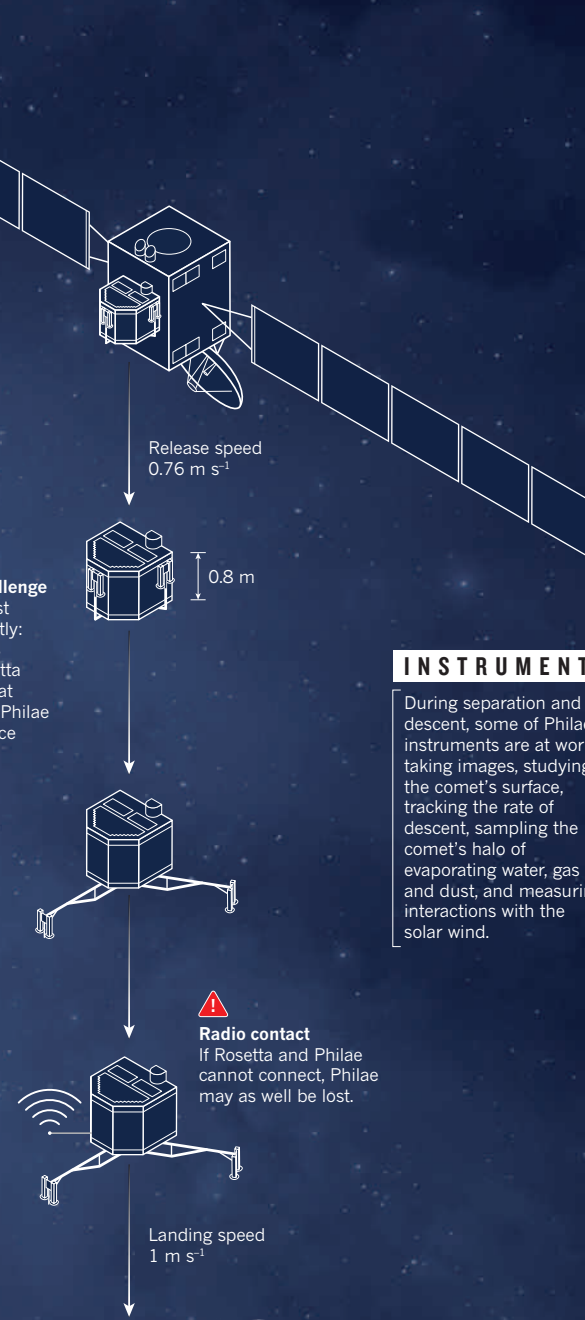


Coordination challenge
The two craft must coordinate perfectly: ejection produces torques that Rosetta compensates for at specific times, so Philae can leave only once Rosetta is ready.

PHONING HOME

T-5h

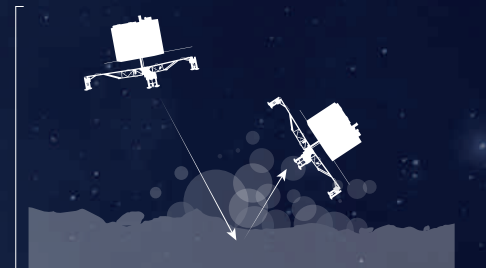
When Philae is about 17 km from the centre of the comet, Rosetta attempts to establish radio contact. Once Philae falls below 10 km, it is in virgin territory. Any data it beams back are completely novel.



TOUCHDOWN

T0h

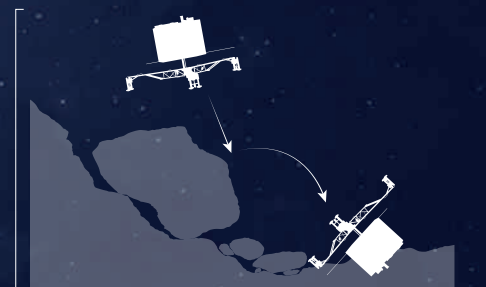
Assuming that mission control got the sums right for the final manoeuvre (see T-9 h), Philae hits somewhere within Agilkia. But much can still go wrong.



The surface is too hard and Philae bounces. The surface is too soft and the lander struggles to secure itself with harpoons and screws. Low gravity makes drilling difficult.

INSTRUMENTS

During separation and descent, some of Philae's instruments are at work, taking images, studying the comet's surface, tracking the rate of descent, sampling the comet's halo of evaporating water, gas and dust, and measuring interactions with the solar wind.



A steep slope or a boulder as small as a chair flips the lander. Philae's legs can cope with a slight tilt, but the lander has no way to fully right itself.

EXPLORATION

Philae's main science phase lasts for three days and includes drilling into the comet's surface and analysing what it finds. But if its solar panels work, Philae can keep going after that, studying how conditions change as the comet gets closer to the Sun and looking for evidence of amino acids. By March 2015, the comet will be too hot for Philae to operate.

Portable chemistry lab, radar, temperature sensors, magnetometer and other instruments probe the comet.

Ice screws
Burrow down to secure Philae.

Harpoons
Fire on landing to anchor Philae.

Drills
Retrieve pristine material from 20 cm down.

Cameras

Take a panoramic picture of the landing site, the first ever from the surface of a comet.



Solar panels

If the climate is dusty, Philae's solar panels might fail, causing the batteries to die after three days. The main science phase would be complete but the end would be premature.

IMAGES: PROBE, ESA/J. HUART; COMET, ESA/ROSETTA/NAVCAM; DESIGN: JASIEK KRZYSZTOFIAK/NATURE

PHILAE ON COMET, ESA/ATG MEDIALAB

Edits to ethics code rankle

Clinical-trial obligations introduced into the Helsinki declaration called too onerous by some.

BY HEIDI LEDFORD

At 50 years old, the rule book that much of the world uses to regulate experiments on humans still stirs debate over what researchers owe to the people who volunteer their bodies to science.

This week, as politicians and ethicists gather to celebrate the anniversary of the World Medical Association's Declaration of Helsinki, drug companies and medical researchers are still grumbling about a pair of revisions made last year. One addressed ethical quandaries raised by a growing trend for clinical trials conducted in the developing world; the other declared that pharmaceutical companies have a responsibility to keep providing medication to participants who benefit from it, even if the drug being tested never makes it to market.

The declaration is a set of principles that many countries draw on when drafting legislation to govern research on human subjects. Institutions often require that their researchers adhere to the declaration, and the World Association of Medical Editors, which includes representatives of 1,000 medical journals in 92 countries, requires studies that have not passed a formal ethics review to follow the declaration's rules as a condition of publication.

"It's very important," says Badri Man Shrestha, a transplant surgeon who conducts clinical trials at the Sheffield Kidney Institute,

UK. "We have to follow every bit of it."

Originally adopted in June 1964, the declaration has remained relevant because it is an evolving document, revised every few years to accommodate trends in medical research. Those changes have also made it controversial; the United States stopped following the declaration in 2008, mainly owing to rules added in 2004 that limited the use of placebos.

Last year's provision on maintaining the supply of a drug after the trial has ended vexes Mark Barnes, a lawyer at Ropes & Gray in Boston, Massachusetts. Known as paragraph 34, it states that trial sponsors, governments and researchers should develop plans before a trial begins, to spell out how they will continue therapy for those people who benefit from an experimental treatment. Barnes worries that this could oblige companies to continue making small quantities of a drug, sometimes at great expense, even if they have decided not to commercialize it. That could siphon off funds that would otherwise be used to develop marketable drugs for larger numbers of patients, he says. And for small companies, Barnes warns, it could spell bankruptcy. "The declaration cannot mean what it says," he adds, "because what it says is absurd."

But others consider post-trial access an

"The declaration cannot mean what it says, because what it says is absurd."

ethical imperative. Gerald Batist, an oncologist at McGill University in Montreal, Canada, once treated a person with pancreatic cancer who responded unusually well to a candidate drug that had otherwise failed in trials. Although pharmaceutical company Bristol-Myers Squibb of New York City ceased development of the drug, Batist and his hospital compelled the firm to continue making it for about seven years. "Obviously that was a significant investment for the company," he says. "But it's a matter of principle." Bristol-Myers Squibb declined to say how much the production run cost.

It was the removal of a provision from the declaration's latest version that rankles Rafael Dal-Ré, a clinical researcher at the Autonomous University of Madrid. He wants to reinstate rules that require trial sponsors to provide "other appropriate care or benefits" — interpreted to include public goods such as schools or medical clinics for communities — when trial participants do not directly benefit from the medicine being tested. This provision, included in the 2008 revision of the declaration, meant that the communities in which trials were conducted, not just the individual research participants, could reap some benefit from their involvement.

The clause, says Dal-Ré, was established in part to compensate for a growing problem: wealthy pharmaceutical companies are increasingly testing drugs in communities where people are generally unable to afford the finished product. In response to the revision, trial sponsors began to build more clinics, schools and clean-water projects in developing countries, he says. But last year, the World Medical Association deleted the provision.

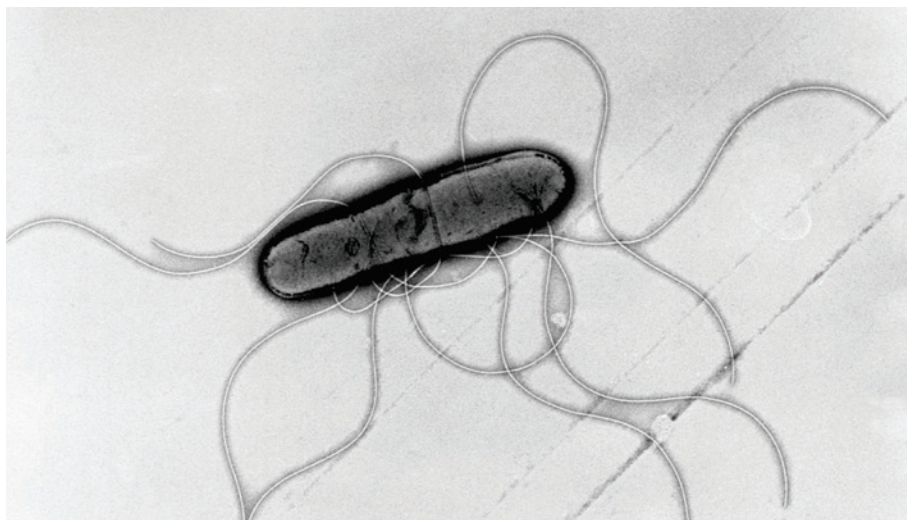
Jeff Blackmer, an executive director of the Canadian Medical Association's ethics office, served on the committee that made the change. He says that the deletion was a response to African physicians who protested that some community leaders were pressuring locals to enrol in clinical trials in order to secure the facilities that came along with them.

Paul Ndebele, director of the Medical Research Council of Zimbabwe in Harare, says that students once contacted him to say that their local leaders were pushing them to enrol in a study. But Ndebele says that the way to surmount that problem is to educate community members, not to remove incentives for helping local residents. "Some of our communities are in pure poverty," he says. "Those communities also need to reap some benefit from these studies." ■



A volunteer in an HIV-vaccine trial in South Africa, subject to the Declaration of Helsinki's ethical rules.

BOSCH/AFP/GETTY



Feeding mice the bacterium *Bacteroides fragilis* can reverse autism-like symptoms.

MICROBIOME

Gut–brain link grabs neuroscientists

Idea that intestinal bacteria affect mental health gains ground.

BY SARA REARDON

Companies selling ‘probiotic’ foods have long claimed that cultivating the right gut bacteria can benefit mental well-being, but neuroscientists have generally been sceptical. Now there is hard evidence linking conditions such as autism and depression to the gut’s microbial residents, known as the microbiome. And neuroscientists are taking notice — not just of the clinical implications but also of what the link could mean for experimental design.

“The field is going to another level of sophistication,” says Sarkis Mazmanian, a microbiologist at the California Institute of Technology in Pasadena. “Hopefully this will shift this image that there’s too much commercial interest and data from too few labs.”

This year, the US National Institute of Mental Health spent more than US\$1 million on a new research programme aimed at the microbiome–brain connection. And on 19 November, neuroscientists will present evidence for the link in a symposium at the annual Society for Neuroscience meeting in Washington DC called ‘Gut Microbes and the Brain: Paradigm Shift in Neuroscience’.

Although correlations have been noted between the composition of the gut microbiome and behavioural conditions, especially autism¹, neuroscientists are only now starting to understand how gut bacteria may influence the brain.

The immune system almost certainly plays a part, Mazmanian says, as does the vagus nerve, which connects the brain to the digestive tract. Bacterial waste products can also influence the brain — for example, at least two types of intestinal bacterium produce the neurotransmitter γ -aminobutyric acid (GABA)².

The microbiome is likely to have its greatest impact on the brain early in life, says pharmacologist John Cryan at University College Cork in Ireland. In a study to be presented at the neuroscience meeting, his group found that mice born by caesarean section, which hosted different microbes from mice born vaginally, were significantly more anxious and had symptoms of depression. The animals’ inability to pick up their mothers’ vaginal microbes during birth — the first bacteria that they would normally encounter — may cause lifelong changes in mental health, he says.

Similarly, a 2013 study from Mazmanian’s lab found that a mouse model with some features of autism had much lower levels of a common gut bacterium called *Bacteroides fragilis* than did normal mice³. The animals were also stressed, antisocial and had gastrointestinal symptoms often seen in autism. Feeding *B. fragilis* to the mice reversed the symptoms. The group also found that the mice with these symptoms had higher levels of a bacterial metabolite called 4-ethylphenylsulphate (4EPS) in their blood, and that injecting that chemical into normal mice caused the same behavioural problems. ►

JOHN MOORE/GETTY

► The mechanism for these effects is still unclear. At the meeting, Mazmanian will present data showing that feeding 4EPS to mice causes behavioural problems only if the gut is leaky, presumably because that allows the chemical to seep into the body through the intestinal wall. That observation raises the possibility that some people with autism could be supported with therapies, such as probiotics, that target the gut instead of the brain, which is a much more complex and inaccessible organ.

Yet even those at the forefront of the research remain sceptical that the findings will translate into treatments for humans. The evidence that probiotics affect human behaviour “is minimal to say the least”, Mazmanian acknowledges. Still, he says, a growing number of researchers are starting to look at some mental illnesses through a microbial lens.

There are implications for basic research too. In another study to be presented at the meeting, veterinarian Catherine Hagan at the University of Missouri in Columbia compared the gut bacteria in laboratory mice of the same genetic strain that had been bought from different vendors. Their commensals differed widely, she found: mice from the Jackson Laboratory in Bar Harbor, Maine, for instance, had fewer bacterial types in their guts than did mice from Harlan Laboratories, which is headquartered in Indianapolis, Indiana.

Such differences could present a major complication for researchers seeking to reproduce another lab's behavioural experiments, Hagan says. When her team transplanted bacteria from female Harlan mice into female Jackson mice, the animals became less anxious and had lower levels of stress-related chemicals in their blood. Hagan notes that when a lab makes a mouse by *in vitro* fertilization, the animal will pick up microbes from its surrogate mother, which might differ greatly from those of its genetic mother. “If we're going to kill animals for research, we want to make sure they're modelling what we think they're modelling,” she says. ■

1. Kang, D.-W. *et al.* *PLoS ONE* **8**, e68322 (2013).
2. Barrett, E. *et al.* *J. Appl. Microbiol.* **113**, 411–417 (2012).
3. Hsiao, E. Y. *et al.* *Cell* **155**, 1451–1463 (2013).



A box of the experimental Ebola treatment ZMapp heads for Liberia.

PUBLIC HEALTH

Ethical dilemma for Ebola trials

Public-health officials split on use of control groups in tests of experimental treatments.

BY ERIKA CHECK HAYDEN

With clinical trials of experimental Ebola treatments set to begin in December, public-health officials face a major ethical quandary: should some participants be placed in a control group that receives only standard symptomatic treatment, despite a mortality rate of around 70% for Ebola in West Africa?

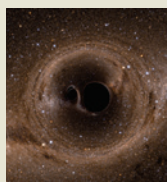
Two groups planning trials in Guinea and Liberia are diverging on this point, and key decisions for both are likely to come this week. US researchers meet on 11 November at the

National Institutes of Health (NIH) in Bethesda, Maryland, to discuss US-government sponsored trials. A separate group is gathering at the World Health Organization (WHO) in Geneva, Switzerland, on 11 and 12 November to confer on both the US effort and trials organized by the WHO with help from African and European researchers and funded by the Wellcome Trust and the European Union.

Experts convened by the WHO in August gave ethics approval to test experimental treatments amid the Ebola epidemic in West Africa. But the WHO trial organizers are not including control arms because most African ►


**MORE
ONLINE**

VIDEO OF THE WEEK



Hollywood and black-hole physics collide (see also Q&A with Kip Thorne page 196)
go.nature.com/d8bjbj

MORE NEWS

- Infiltrator of blood-brain barrier targets Alzheimer's go.nature.com/qfoofk
- Half of the Universe's stars could lie outside galaxies go.nature.com/kfmptr
- Feeling ghostly presence is just brain misintegrating signals go.nature.com/jtaslp

NATURE PODCAST



The depths of depression; best diets for the planet; and more on the movie *Interstellar*
nature.com/nature/podcast

THE GREAT DEPRESSION

DEPRESSION CAUSES MORE DISABILITY THAN ANY OTHER DISORDER. A SPECIAL ISSUE EXPLORES HOW SCIENCE CAN HELP.

A few months after the world went grey, Sue Wright checked into a hospital. A social slight had flipped a switch in her mind, draining life of colour and joy. Blue skies became dull; laughter was unthinkable. Often, the depression left her bedridden. “I had prided myself on being able to get through anything,” says Wright, now a social worker in Germantown, Maryland. “Suddenly, I couldn’t.”

Wright’s story is familiar to too many people. Depression is not just the most common mental-health disorder: it is responsible for a greater burden of disability than any other cause. In this special issue, *Nature* asks why that burden is so great, how science is helping and where research is running aground.

A graphic tour on page 180 shows that depression is far from a Western blight, and that many of the countries most afflicted by it are those with the least resources to help. Some mental-health experts say that the high levels of undiagnosed or untreated depression would not be tolerated for a disease such as cancer, and a News Feature (page 182) examines this claim. It

finds that the absence of a crisp diagnosis and a lack of tools to understand the brain’s complexities have held back therapy and research.

The urgent question is how to overcome those barriers, and scientists are exploring several routes. Some argue that there is much to be learned from studying the mechanisms of existing antidepressants; others that there is most promise in teasing apart the affected brain circuits (see page 200) or gleaning information on common medicines that might have unexpected benefits for brain disorders (see page 165). Identifying the genes associated with depression has been a thankless task, but ambitious studies involving many thousands of patients are now called for (see page 189). There is also plenty to be done to refine existing treatments, such as cognitive behavioural therapy, and to tailor them to groups who might benefit most (page 185).

Medication, counselling and electroshock therapy did not work for Wright. After trawling through medical journals, she found a psychiatrist prescribing drug combinations that may boost the effect of antidepressants. After weeks of one such combination plus therapy, Wright realized that the sky was blue again. It took months for her to find a way out; the hope is that research will find a faster route to relief. ■

DEPRESSION

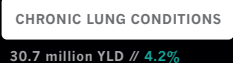
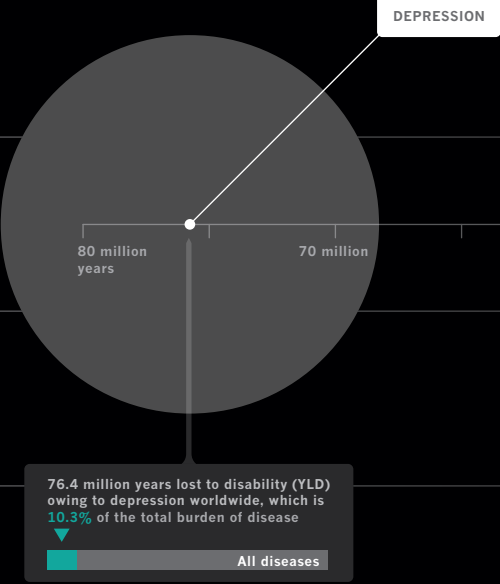
A *Nature* special issue
nature.com/depression

ILLUSTRATION BY CARL DE TORRES

A world of depression

Depression is a major human blight. Globally, it is responsible for more ‘years lost’ to disability than any other condition. This is largely because so many people suffer from it — some 350 million, according to the World Health Organization — and the fact that it lasts for many years. (When ranked by disability and death combined, depression comes ninth behind prolific killers such as heart disease, stroke and HIV.) Yet depression is widely undiagnosed and untreated because of stigma, lack of effective therapies and inadequate mental-health resources. Almost half of the world’s population lives in a country with only two psychiatrists per 100,000 people.

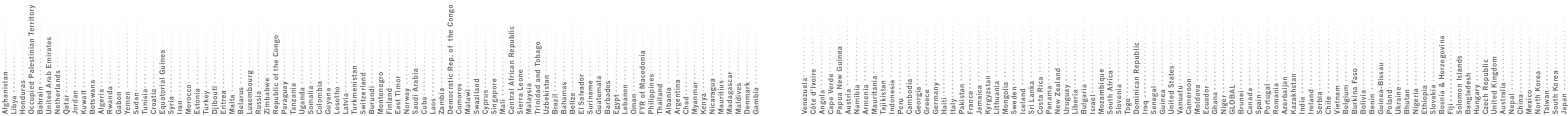
● Prevalence of depression ● Psychiatrists per 100,000 people



Top ten causes of disability

Depression accounts for the biggest share of the world’s burden of disease, measured by years lost to disability (YLD): healthy years ‘lost’ because they are lived with a physical or mental disability.

● Percentage of total burden of disease



PSYCHIATRISTS PER 100,000 PEOPLE

NO DATA AVAILABLE

Afghanistan ● 22.5% ● 0.16

This war-torn country reports the world’s highest prevalence of depression, and yet it is among the least equipped to deal with it. Conflict is a well-established risk factor for depression, as are child sexual abuse and domestic violence.

Switzerland ● 6.16% ● 41.42

Switzerland’s universal health care offers some of the best support for mental health, with more than 40 psychiatrists per 100,000 people. But studies estimate that even the best available treatments can reduce the burden of disability from depression by only 10–30%.

United States ● 4.45% ● 7.79

Depression receives significantly less in research funds from the US National Institutes of Health than do cancer or heart disease. That is partly because of a lack of patient advocates and the stigma that surrounds the condition.



China ● 3.02% ● 1.53

The relatively low prevalence of depression in China could be the result of the way in which it is diagnosed, rather than lower actual rates. People with depression here often report symptoms such as stomach pain or headache, and so may be missed by the standard diagnostic criteria used to calculate prevalence, which focus on mood, lack of motivation and fatigue.

BY KERRI SMITH | INFOGRAPHIC BY CARL DE TORRES



SYLVAIN CHERKAOU/COSMOS/EYEVINE

A nurse attends to a patient at a Médecins Sans Frontières Ebola treatment centre in Kailahun, Sierra Leone.

Tough choices to reduce Ebola transmission

Christopher J. M. Whitty and colleagues explain why the United Kingdom is funding many small community centres to isolate suspected cases in Sierra Leone.

The Ebola outbreak in West Africa is larger than all previous Ebola outbreaks combined, and is still expanding¹. With a death toll in the thousands, and mortality at around 70%, it has undermined fragile health-care systems by filling hospitals with highly infectious patients and killing health workers.

Transmission in dense urban populations presents challenges never before seen. Vaccines and drugs would offer a complementary approach to control and they must be fast tracked. But these medicines are not available for immediate widespread use. To quell this unprecedented outbreak, we have to use methods that have not been tested.

Experience from past outbreaks has established reliable methods to control transmission in hospitals and at funerals of people who die from Ebola — two of the three main venues for transmission (see 'Hospitals and funerals'). These efforts remain essential, and the scientific and operational strands must complement one another.

What most distinguishes the current situation from previous outbreaks is the high proportion of transmission occurring in the community. No techniques have been developed to control community transmission at this scale^{2,3}. Notwithstanding uncertainties, the global community must act now.

The UK government is leading the international response to Ebola in Sierra Leone, providing technical, financial and logistical help. This article sets out the scientific basis for the UK government's strategy to assist Sierra Leone's government to reduce transmission. In addition to substantially scaling up conventional capacities at hospitals, we plan to help to build and support community isolation centres where people can voluntarily come to be isolated if they suspect that they have the disease.

These interventions represent uncertain ground, but hesitation is more dangerous than trying out potentially ineffective methods (see 'Delays mean more deaths'). Climbing rates of infection could soon overwhelm

Sierra Leone's already challenged health services, and so remove any realistic chance of public-health intervention ending the outbreak. The Sierra Leonean health system and the many brave local and international health staff need help to lower transmission to the point at which mopping up remaining islands of high transmission will be possible.

Until the outbreak is over, we will not know whether we have launched the best response. We invite critiques and suggestions, but must act swiftly. Further delay will result in more infections and deaths, and only sabotage future efforts.

THE CHALLENGE

About 500 new cases of Ebola are being diagnosed every week. This number is expected to double in the next 30 days⁴. The force of transmission of a disease outbreak is quantified as R , the average number of people infected by each newly infected person (the reproduction number). When $R = 1$, an outbreak is neither increasing nor decreasing.

EBOLA TRANSMISSION

Hospitals and funerals

The initial symptoms of Ebola (which may include fever, diarrhoea and vomiting) mimic those of many common diseases, including malaria, pneumonia and gastrointestinal infections. Transmission occurs through bodily fluids: diarrhoea, vomit, blood and probably sweat and semen. Patients become infectious when they are symptomatic, and remain highly infectious until they begin to recover. Corpses are highly infectious.

Controlling transmission requires minimizing contact with bodies, bodily fluids and contaminated items such as mattresses, clothes and clinical waste. Transmission does not occur through breathing or insects. Transmissibility through existing modes of infection could evolve as the outbreak expands.

Ending previous Ebola outbreaks relied on stopping transmission in hospitals through strict infection control; adapting funeral rites to reduce risk of transmission; helping local people to protect themselves; and actively tracking down those likely to be infected⁷.

Infection control in hospitals requires sufficient resources and trained personnel to implement measures, such as using personal protective equipment. In relatively well-equipped hospitals with stringent training and control protocols (such as those run by Médecins Sans Frontières), infection rates remain low.

Reducing transmission from the dead is more complex, and social scientists play an essential part. Protocols for handling and

disposing of a body to minimize the risk of infection are well established, but must be implemented with the cooperation and consent of grieving communities. Funeral rites often require that family members and mourners have extensive contact with the deceased.

Previous Ebola outbreaks in the Democratic Republic of Congo and Uganda have shown that unless infection-control routines can be adapted to local burial practices, communities will continue their traditions. Anthropologists have documented funeral rites in Sierra Leone and other countries. Dialogues with Ebola-stricken communities are essential to adapting their traditions to reduce body handling respectfully.

According to our analyses, R in Sierra Leone is currently between 1.2 and 1.5. In some areas, R is considerably higher. If R remains above 1, any public-health intervention (except a vaccine) will eventually be overwhelmed by the number of new infections. Getting R below 1 is the single strategic aim of the UK effort at this stage of the outbreak.

Reducing R depends on stalling transmission in the community, which depends on reducing the time between when people first show symptoms and are isolated. Delay occurs in two broad settings: patients who have symptoms consistent with Ebola but have not been positively diagnosed; and people diagnosed with Ebola who are being nursed at home because of reticence to move to facilities, or a lack of available health-care beds or transport. Patients in the later stages of Ebola are the most infectious, in part because symptoms such as vomiting and diarrhoea are worse⁵. Models vary, but most work, including that led by N.F. and W.J.E., suggests that isolating 70% or more of Ebola cases within three days of their becoming symptomatic is key to reducing R to 1.

TWO-PRONGED STRATEGY

Designing measures to control community transmission poses epidemiological, clinical and social-science challenges. Currently, most people with early Ebola-like symptoms will not have Ebola, but diseases such as malaria and pneumonia. Undiagnosed individuals should be isolated both from health workers (as in Ebola wards) and from each other, so that the true Ebola cases do not infect those with other diseases. The high number of suspected cases greatly complicates isolation efforts.

Community control of previous Ebola

outbreaks relied on active 'case-finding': tracing contacts of known cases and then monitoring and quickly isolating suspected cases early. It can take as many as ten staff to monitor the contacts of just a single case, and the task is potentially more difficult in mobile urban populations, especially as Ebola case numbers rise.

The ideal approach would be active case-finding combined with isolating patients in fully equipped and staffed hospitals, but this is not practical in the current situation.

"Centres must be able to operate with small, relatively unspecialized staff drawn from the community."

In some areas, the outbreak has already overrun hospitals; many suspected Ebola cases are being turned away for lack of beds. An unknown proportion of cases remain at home. For those who seek medical care, the current median time between becoming symptomatic (and thus infectious) and isolation in Sierra Leone is four days. Many wait more than a week.

These delays will only get longer. Current UK aid efforts — which have so far pledged more than £200 million (US\$320 million) and the largest UK troop deployment outside Afghanistan — will help to increase Sierra Leone's bed capacity threefold by January. But at current measures of R , the projected increase in new cases (thousands per week) will far exceed the number of possible new hospital beds. To avoid that scenario, beginning this month, affected regions must substantially increase rates of early isolation for suspect and confirmed Ebola cases.

One proposed strategy — giving families information and basic personal protective

equipment (PPE) to minimize transmission while nursing patients at home — is problematic. Using PPE safely is difficult even for professionals, as infection rates in health-care workers demonstrate. And identifying cases and training families requires staff that Sierra Leone does not have. This approach is acceptable only as a desperate humanitarian measure when there is no space available in health facilities. It is not a good strategy to reduce transmission.

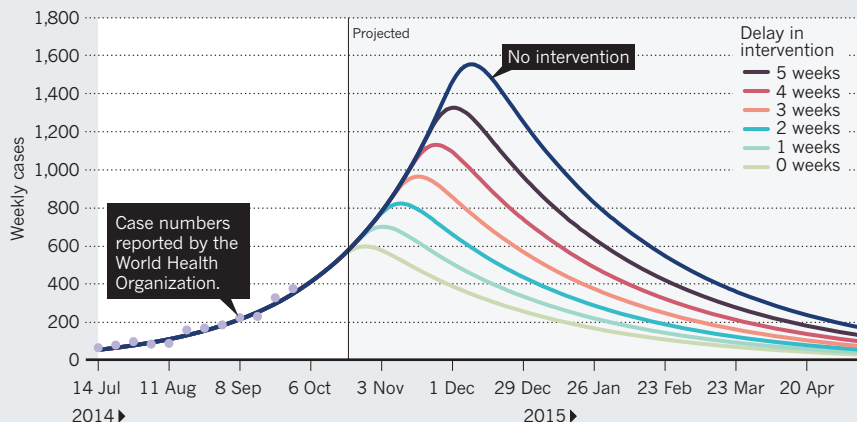
PASSIVE CASE-FINDING

The UK government has decided to support another strategy: passive case-finding with community isolation. This is at the request of the government of Sierra Leone and endorsed by the World Health Organization. The strategy, also called voluntary sequestration, encourages those with suspected Ebola infection to gather in units where they can receive basic health care and avoid infecting members of their families and communities. By quickly isolating suspected cases (and boosting the proportion of such cases isolated early), this approach aims to keep people with Ebola from infecting others, until R is pulled below 1.

Some models are currently being piloted near Sierra Leone's capital, Freetown, co-designed with local clinicians and communities. An initial 200 community units can be expanded according to how and where the outbreak progresses. All share similar requirements. Community-level units must be easy to construct, relying mainly on tents or repurposed buildings in or near affected communities, that can be set up in days or weeks. Many small units are better than a few large ones. They can be placed closer to communities (distance can be a barrier to people

DELAYS MEAN MORE DEATHS

For every week that effective interventions are not implemented in Ebola-stricken areas, the case-number peak will be pushed higher and later, meaning that more hospital beds will be needed.



checking themselves in). It also reduces the risk of a major outbreak if infection control breaks down in one unit. Centres must be able to operate with small, relatively unspecialized staff drawn from the community, supported with basic training and provision of equipment and supplies by agencies such as non-governmental organizations (NGOs).

Most importantly, centres must not become sources of new infections. This means effective isolation, as well as diagnosing and moving confirmed Ebola cases on to specialized treatment facilities as quickly as possible.

Because most patients in the early stages of the disease will have relatively minor symptoms, one health worker can care for several people. And because patients choose to report to these centres, communities are more likely to accept them. Even incomplete coverage could yield successes. The aim is not to find every suspected case, but to attract enough early cases to shorten the time that infectious people spend in the community, and so reduce R .

DATA, MODELLING AND RISK

There is little current evidence from previous outbreaks on how well, or even whether, voluntary community isolation reduces transmission. We will need to learn and adjust as data come in. Accurate, recent, open data are essential to shape an effective response, but data-capture systems were implemented only in September in the worst-affected countries. These now allow us to estimate R more accurately and to track time between symptoms and isolation.

Major uncertainties remain — most notably around the proportion of Ebola cases currently being detected by health systems. If only a small proportion of cases are being identified, then a dramatic increase in self-reporting is essential. Conversely, if the major issue is, as we suspect, delays in isolation, then even a one- or

two-day reduction in that time might be enough to achieve control.

Modelling can help to predict what provisions are needed. Our current projections of the outbreak in Sierra Leone suggest that more than 200 ten-bed isolation units will be required by the end of December. This estimate may be too low.

The UK government has been incorporating social-science advice into the incentives for bringing sick patients to community centres⁶. The desire for good medical care, for protecting loved ones from infection and possibly provision of food or money would all encourage people with early symptoms to go to Ebola evaluation centres. But distance, suspicion, fear and stigma will discourage early reporting. Under-incentivizing would mean that too few people with Ebola will report to centres to reduce transmission. Over-incentivizing could mean that too many non-infected people report, overwhelming facilities and increasing risks of transmission. To achieve the optimal balance, facility planners will need to engage with local populations and be flexible enough to adopt better practices as evidence accumulates.

Careful collaborative work with communities, building on what is already known from anthropological studies and field reconnaissance, must guide how facilities are located, designed, staffed and operated to maximize social acceptance.

This public buy-in is essential. Communities have reacted negatively to the opening of treatment facilities, and in some cases, people have forcibly removed family members from isolation units. Rumours of nefarious killing and body-part theft have circulated, reflecting anxieties that resonate with people's experiences of disease, war and government intervention (see www.ebola-anthropology.net). The murder of health workers and journalists in Guinea this September exemplifies the risks to staff if interventions go wrong.

Optimal design will be learned on the fly. We know how to minimize infection in dedicated Ebola wards with highly trained and supervised staff, but not in facilities with lightly trained staff in which most people do not have Ebola (but some do). Evidence will have to be collected on the best ways to design units, place patients, speed Ebola diagnoses and transport cases. There are likely to be early failures from which the global community must learn. Given the risks that health workers take, getting Ebola vaccines to staff working in the units as soon as they have proved safe is an ethical imperative.

Extrapolating from data collected in neighbouring countries on common pathogens and antimicrobial drug resistance will be essential. Diagnostic testing, except for Ebola, must be kept to a minimum to avoid contamination from blood and sharp instruments. Health workers should use treatments on the basis of symptoms alone to deliver appropriate care for malaria and bacterial disease without testing, a practice that has been used effectively in many under-resourced settings. Faster and safer Ebola tests (those that would not require using blood with the live virus) would substantially improve clinical management.

We hope that in the coming months, safe and effective vaccines will be produced at large enough volumes to stop this outbreak, but thousands are dying now. Qualitative, quantitative and clinical sciences, and the ability to adapt and learn from mistakes, are urgently needed to reduce community transmission. ■

Christopher J. M. Whitty is chief scientific adviser at the UK Department for International Development. **Jeremy Farrar** is director of the Wellcome Trust in London, UK. **Neil Ferguson** is professor of mathematical biology at Imperial College London, UK. **W. John Edmunds** is professor of infectious-disease modelling at the London School of Hygiene & Tropical Medicine, UK. **Peter Piot** is director of the London School of Hygiene & Tropical Medicine, UK. **Melissa Leach** is director of the Institute for Development Studies in Brighton, UK. **Sally C. Davies** is chief medical officer and chief scientific adviser at the UK Department of Health. e-mail: c-whitty@dfid.gov.uk

1. WHO Ebola Response Team *N. Engl. J. Med.* **371**, 1481–1495 (2014).
2. World Health Organization Ebola Response Roadmap Situation Report: 31 October 2014 (WHO, 2014).
3. Kerstiens, B. & Matthys, F. *J. Infect. Dis.* **179** (Suppl. 1), S263–S267 (1999).
4. Borchert, M. et al. *BMC Infect. Dis.* **11**, 357 (2011).
5. Bausch, D. G. et al. *J. Infect. Dis.* **196** (Suppl. 2), S142–S147 (2007).
6. Hewlett, B. S., Epelboin, A., Hewlett, B. L. & Formenty, P. *Bull. Soc. Pathol. Exot.* **98**, 230–236 (2005).
7. Heymann, D. L. *Nature* **514**, 299–300 (2014).



Left, Benedict Cumberbatch as Alan Turing in *The Imitation Game*; right, Eddie Redmayne as Stephen Hawking in *The Theory of Everything*.

FILM

Enigma variations

Robert P. Crease ponders a brace of biopics on Alan Turing and Stephen Hawking.

Early in *The Imitation Game*, a constable says, “I think Alan Turing’s hiding something.” The implication is that it is his sexuality. But as well as being gay, Turing was a master code-breaker, spy and recluse. Secrets drive this thrilling, fast-paced film — as gripping as it is tragic — about one of history’s most brilliant mathematicians and computer engineers. Meanwhile, a movie about physicist Stephen Hawking turns his life into a feel-good saga.

Morten Tyldum’s *The Imitation Game* makes it clear that Turing’s passion was finding solutions to abstract logical problems. But how to animate a character (played by Benedict Cumberbatch) whose life is all in his head? The answer is provided by the Enigma machine, used by Nazi Germany to encipher military messages. In trying to crack the secret, Turing’s otherworldly interests in ‘impossible’ logical challenges intersected with urgent worldly ones — the challenges of the British government, at one of the darkest junctures of the Second World War.

An early scene, in which we are first shown Enigma, dramatizes that intersection. Turing pronounces the device beautiful; a government official declares it “the crooked hand

The Imitation Game

DIRECTED BY: MORTEN TYLDUM
Black Bear Pictures: 2014.

The Theory of Everything

DIRECTED BY: JAMES MARSH
Working Title Films: 2014.

of the devil”. As Turing and his colleagues at Bletchley Park, the British code-breaking centre, work on deciphering Enigma, every tick of the clock means more warfare, more death. The film cross-cuts scenes of bombs raining on cities with scenes of prospective code-breakers solving crossword puzzles, and you get the connection.

When the Bletchley gang finally solve the code, they have to hide their triumph or risk tipping the enemy off. Turing is asked to determine which secret — such as news of impending attacks on convoys or cities — can be acted on without arousing suspicion. Meanwhile, the Soviets are spying on Bletchley, kept at bay with a drip-feed of minor information. The UK government blackmails Turing (gay in an era when that was illegal) to force him to spy on the Soviet ring.

The film takes its title from the Turing test, an imitation game that he devised in which a human has to judge whether a concealed interlocutor is a machine or a human. The idea is invoked in a scene, played powerfully by Cumberbatch, in which Turing is about to reveal the full details of his life story, including the code-breaking and homosexuality. He challenges the (baffled) constable to judge him: “What am I? Am I a criminal? Am I a war hero?”

The movie neither sensationalizes nor ignores Turing’s sexuality. It tells us little about his work in maths or cryptography, but it allows us to glimpse something deeper: how the ordinary world is shaped by mathematicians and scientists in ways that many people hardly suspect. A final scene shows ecstatic celebrations in the street at war’s end. We are left feeling that, if such events fail to give the Turings of the world their due, the revelry is hollow. Turing, of course, was not just unheralded: like most Bletchley employees, his accomplishments were classified and his war records expunged. Later, he was arrested for ‘gross indecency’ and forced to undergo chemical castration. He killed himself in 1954.

► **NATURE.COM**
For Nature’s Alan Turing special, see: nature.com/turing

► *The Theory of Everything* is, like *The Imitation Game*, about a scientist. There the similarities end. The film tells us that Hawking has refused a knighthood, but it thoroughly canonizes him. There have been other depictions of Hawking's life, including *Hawking* (2013), a documentary by Stephen Finnigan, and *Hawking* (2004), a BBC television film starring Cumberbatch. This turn, with Eddie Redmayne in the lead role and directed by James Marsh, is loosely organized around Hawking's relationship with his first wife, Jane Wilde, and based on her memoir *Travelling to Infinity* (Alma, 2008). Despite Redmayne's fine performance, it unfolds with such plodding, reverential linearity, and with so much melodrama and uplift, that it is the least compelling of the treatments.

The story begins in Cambridge, UK, in 1963. Stephen and Jane flirt. He is diagnosed with motor neuron disease and told that he has two years to live. She remains selflessly devoted and they marry. Stephen works despite his disability. "He has done it!" we are told at one point, but we haven't a clue as to what 'it' is. Caring for Stephen and their three children gives Jane moments of doubt and insecurity. The ins and outs of the couple's relationships with each other and with their eventual second spouses — choir director Jonathan Hellyer Jones and Stephen's nurse Elaine Mason — are rolled out with surprisingly little conflict. Everyone is presented as well intentioned and concerned only for Stephen's welfare.

Key moments in Hawking's life are handled simplistically. He announces, "I will write a book," and knocks off *A Brief History of Time* (Bantam, 1988) without strain or editorial conflict; it is an instant best-seller. A final scene shows Hawking, now a venerated celebrity, delivering a talk to an entranced audience. One questioner asks how it is possible for him to live without believing in God. The question seems pointless: in the world of this film, so full of saints and sacrifice, who needs God? But cloaked in the ethereal strains of his electronically synthesized voice, Hawking's response — "Where there is life, there is hope" — brings the audience cheering to its feet. The scene invites us to regard Hawking himself as a miracle.

Dramatizing the lives of scientists, so much of whose work is mentation, is demanding. These two films demonstrate the promise and the perils. ■

Robert P. Crease is professor of philosophy at Stony Brook University in New York, and co-author, with Alfred Scharff Goldhaber, of *The Quantum Moment*.
e-mail: robert.crease@stonybrook.edu



Matthew McConaughey, Anne Hathaway and David Gyasi in *Interstellar*.

Q&A Kip Thorne

Space-time visionary

Thanks to theoretical physicist Kip Thorne, real science is embedded in Christopher Nolan's film *Interstellar*, in which explorers seek a new home for humankind. Thorne talks about what he learned from the film's unprecedented visualizations of black holes and wormholes, what it and his accompanying book can teach, and the likelihood of humans escaping the Solar System.



How did *Interstellar* come about?

I have long worked on black holes and, since the 1980s, wormholes — hypothetical tunnels in space that link distant regions of the Universe. About eight years ago, I and my

friend Lynda Obst, a film producer, came up with a movie set on the 'warped side of the Universe' — black holes, wormholes, higher dimensions and beyond. It interested director Steven Spielberg, who brought in Jonathan 'Jonah' Nolan to write the screenplay. Steven dropped out and later Jonah's brother Christopher Nolan took over as director and final screenwriter. Chris and Jonah changed our story almost completely, but preserved the warped space-time and splendidly fulfilled our vision of a science-fiction movie with real science woven deeply in its fabric. In it, nothing violates well-established physical laws and all the wild speculations spring from science, not just the fertile mind of a screenwriter.

How hands-on were you during development?

I met with Jonah and Chris every few weeks as they crafted the screenplay, brainstorming

Interstellar

DIRECTED BY CHRISTOPHER NOLAN
Warner Brothers: 2014.

The Science of Interstellar

KIP THORNE
W. W. Norton: 2014.

about the science. I worked on the visualization of black holes and wormholes with Oscar-winner Paul Franklin and his team at Double Negative Visual Effects in London.

Black holes do not emit light, so you visualize them through gravitational lensing — how they bend light from other objects. I took equations based on Einstein's general theory of relativity and created a description of a wormhole with three parameters: diameter, interior length and the degree of flare where the wormhole joins the external Universe. Paul's team used my equations to compute what a camera would see through the wormhole; Chris, perusing the images, chose the parameter values for *Interstellar's* wormhole.

Did you learn anything new?

With computing power beyond what is normally used by physicists, and software designed to give rapidly changing images at IMAX resolution, we were able to see something never seen before. We simulated a fast-spinning black hole and a field of stars, and what we discovered is an amazingly complex,

ILLUSTRATION BY JIM SPENCER

Correspondence

Let poor countries into rich research

Researchers and decision-makers from low-income countries should help to define the research priorities of universities in high-income countries (see *Nature* **514**, 287; 2014). A strong guild of academics will aid the development of local solutions in poor countries that have little resilience to climate change, biodiversity loss or global health issues (see, for instance, C. Lachat *et al.* *PLoS Med.* **11**, e1001593; 2014).

It is all too common for Western scientists to engage in research in the developing world that fails to benefit their locally based peers (see K. Chu *et al.* *PLoS Med.* **11**, e1001612; 2014). This inequity could be rectified by introducing and enforcing strict ethical standards for funding and research. More-equitable metrics for research evaluation should also be developed (see, for example, E. Smith *et al.* *BMC Med. Ethics* **15**, 42; 2014).

Carl Lachat Ghent University, Belgium.

Patrick Kolsteren Ghent University, Belgium; and Institute for Tropical Medicine, Antwerp, Belgium.

Dominique Roberfroid Institute for Tropical Medicine, Antwerp, Belgium.

carl.lachat@ugent.be

Better modelling for the energy mix

Contrary to Haewon McJeon and colleagues' implications (*Nature* **514**, 482–485; 2014), natural shale gas and renewables are not necessarily competitors. They are driven by different policies and markets.

Over the next decade or so, natural gas can reduce coal consumption and can be rapidly dispatched to balance the intermittency of energy from wind and solar sources. In the longer term, say over the next

50 years, this mix can provide a path to an affordable, low-carbon energy system, especially if it is configured as a backup or coupled with energy storage.

The low capital and high operating costs for natural gas combine well with the high capital and low operating costs of wind and solar. And by reducing the cost of climate mitigation, natural gas should motivate more-aggressive climate policies by governments. Gas is therefore promising because it is a powerful climate-mitigation option, not because it will prevent climate change in the absence of policy.

Moreover, the decisions that shape global energy choices are not always economic or technical (see B. K. Sovacool *Nature* **511**, 529–530; 2014). Effective modelling needs to capture all of the complex and often non-quantitative forces that determine outcomes. These include energy security, trade policy, social equity, weather, distribution of resources, dynamic policies and unforeseen technological advances. Without these factors, the reliability of the long-term forecasts from McJeon *et al.* seems low.

Alexander Q. Gilbert Haynes and Boone, Washington DC, USA.

Benjamin K. Sovacool Aarhus University, Denmark.
benjaminso@hih.au.dk

Funding plea for rural lab outreach

The Imagine Project, set up last year by a group of Brazilian researchers (see go.nature.com/yndqs1), is taking scientific knowledge out of the laboratory to share it with people from rural and indigenous communities. Despite their remarkable response, the initiative is still struggling to be funded.

Scientists and students from universities in Brazil, Morocco, Angola, Mexico and Peru are teaching, for example, molecular biology to Brazilian Indians, genome analysis to Peruvian Inca

descendants and facts about solar energy to teenage land-workers (see go.nature.com/fdqnws).

Community members work and study on alternate days. The learners have, without exception, shown great enthusiasm. Their manual skills often exceed those of urban university students.

The project has been totally funded by Brazil's Federal University of Santa Catarina since its inception. Applications to corporate and governmental funding organizations have so far met with a dispiriting response.

The knowledge imparted is viewed as “too universal” — with no potential for immediate conversion into solutions for food, health or environmental problems. Policy-makers seem to regard the people who are engaged as too remote, too few and too ‘invisible’ to justify involving them in modern developments in science and technology.

André Ramos, Guilherme Razzera Federal University of Santa Catarina, Florianópolis, Brazil.

andre.amos@ufsc.br

Approximation aids handling of big data

We need a radical shift in our approach to data analysis — towards approximation. Our technical capacity is being overtaken by the unprecedented rate of data generation by today's powerful instruments and computers (see H. Esmaeilzadeh *et al.* *Commun. ACM* **56**, 93–102; 2013).

In my view, it is often more efficient to replace conventional, precise computations with scalable approximations that have tight error bounds — in algorithms and data structures, for example. This should be adequate for interpreting data, particularly when refining the explorative phase of an analysis. Researchers may ultimately

still want to perform detailed analyses as indicated trends become clearer. And for some applications, approximation will not work at all.

Approximations, however, are already accepted in many realms, such as next-generation DNA sequencing (M. L. Metzker *Nature Rev. Genet.* **11**, 31–46; 2010). It is time to let mathematical ingenuity replace our obsession with precision.

Thomas Heinis Imperial College London, UK.

t.heinis@imperial.ac.uk

EC grant applicants need fuller feedback

The commitment of the European Commission (EC) to equitable and rigorous scientific review of the research it funds is undermined by the perfunctory quality of feedback to grant applicants. This denies investigators the opportunity to learn and improve from constructive expert criticism.

The commission's review procedures are time-consuming and costly. So, regardless of the final decision on funding, individual applicants should be privy to the full text of expert critiques of their proposals — not merely to a brief summary of the consensus ruling. The anonymous reviewers also deserve to be heard.

In my experience, these evaluation summary reports contain only vague comments, without constructive suggestions or scientific advice. For grant applications that can run to 70 pages, this minimal feedback seems inadequate.

The European Union cannot afford the current waste of intellectual resources and input into scientific development.

Danuta Wasserman European Psychiatric Association, Strasbourg, France; and Karolinska Institute, Stockholm, Sweden.

danuta.wasserman@ki.se

RESEARCH INTO DEPRESSION HAS STRUGGLED, WHILE STUDIES OF CANCER HAVE THRIVED — BUT THE BALANCE COULD BE SHIFTING.

IF DEPRESSION WERE CANCER

BY HEIDI LEDFORD

If the extent of human suffering were used to decide which diseases deserve the most medical attention, then depression would be near the top of the list. More than 350 million people are affected by depression, making it one of the most common disorders in the world. It is the biggest cause of disability, and as many as two-thirds of those who commit suicide have the condition.

But although depression is common, it is often ignored. Three-quarters of people with depression in the United Kingdom go undiagnosed or untreated — and even if the disorder is diagnosed, today's medications will work well for only about half of those who seek help. "It's unbelievable," says Tom Foley, a psychiatrist at Newcastle University, UK. "If that was the case in cancer care, it would be an absolute scandal."

The comparison between depression and cancer is a common one. Cancer, too, is a terrible blight: it affects more than 32 million people and kills some 8 million a year, many more than depression. But at least in developed countries, the vast majority of people with recognized cancers do receive treatment.

In research, too, depression has failed to keep up with cancer. Cancer research today is a thriving field, unearthing vast catalogues of disease-associated mutations, cranking out genetically targeted therapies and developing

sophisticated animal models. Research into depression, meanwhile, seems to have floundered: once-hopeful therapies have failed in clinical trials, genetic studies have come up empty-handed. The field is still struggling to even define the disease — and overcome the stigma associated with it.

Depression research also gets a great deal less funding than that gobbled up by cancer. The US National Institutes of Health pumped about US\$5.3 billion into cancer research in 2013 — a stark contrast to the \$415 million it spent on depression research and the \$2.2 billion on mental-health research as a whole. The same pattern holds elsewhere: in its most recently completed funding scheme, the European Union invested about €54.3 million (US\$67.4 million) a year for studies of mental-health disorders, €8 million of which was flagged specifically for depression. The programme allotted €205 million a year for studies of cancer.

No one denies that cancer deserves rich funding and attention, nor do they begrudge the advances made in understanding the disease. Mental-health researchers just wish that they could claim similar advances for their

field, and that medical care could offer more.

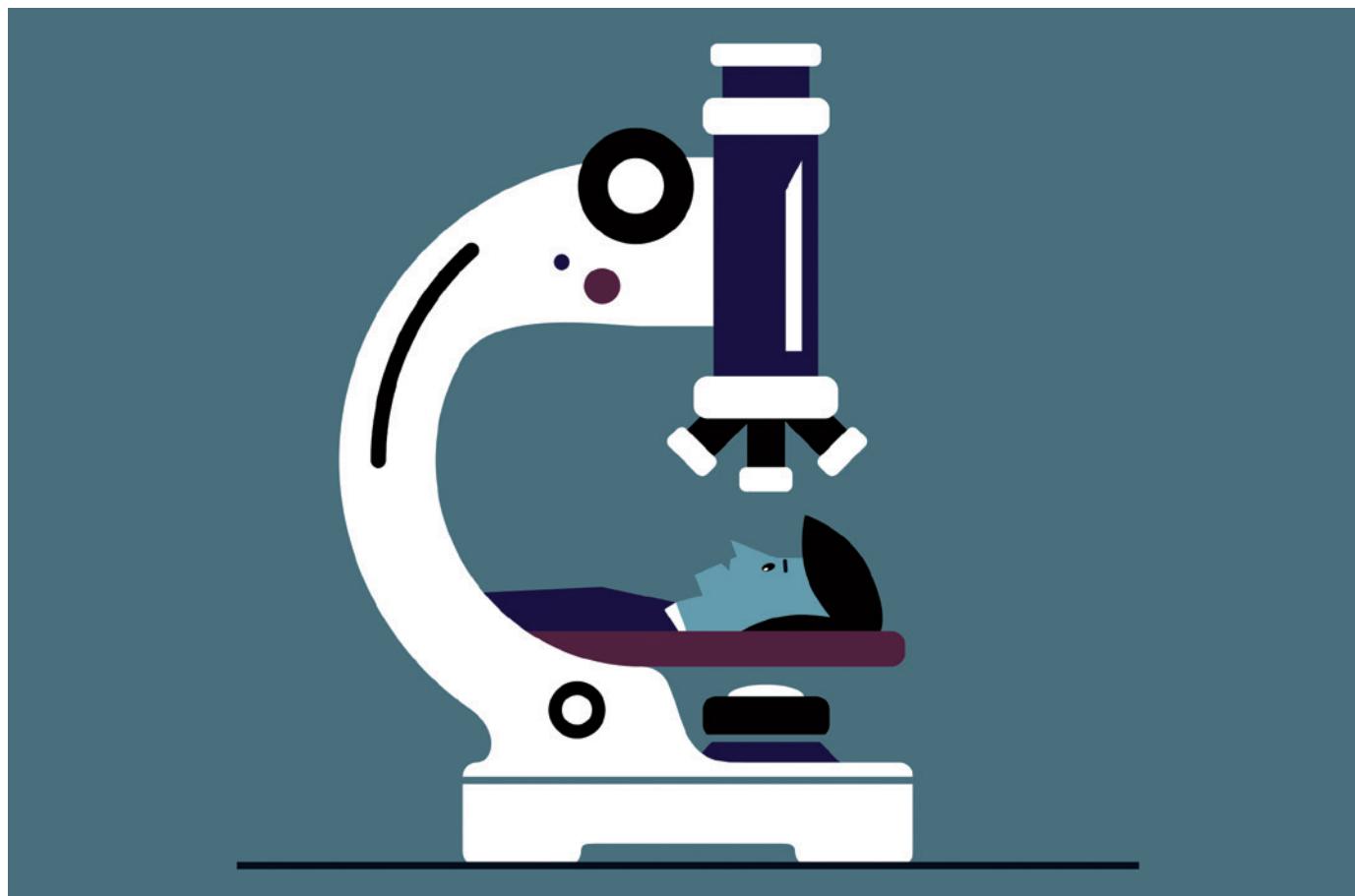
So why has depression not garnered the same scientific resources and attention as cancer? And had it done so, where would understanding of this disorder stand now? *Nature* put these questions to researchers. Although many said that extra money would have solved some challenges earlier, the technology needed to crack others — by probing the brain and analysing its circuits, for example — is only now emerging. But some scientists hope that a recent explosion of interest in brain studies will at last push mental-health research into a different league. "Cancer's a great inspiration: they've had a lot of investment and they've made big breakthroughs," says Foley. "There's no reason why we can't see the same things in depression."

POWER OF ADVOCACY

Research agendas are rarely set by human need alone. Political, social and economic concerns can all tip the balance in favour of one disease or another — and patient advocates have a major influence on the way that money is handed out. The divide between cancer and depression can be traced back several decades, when strong advocacy helped to spur the United States to declare a 'war on cancer' in 1971. Since then, funding has poured into the field, seeding a huge research enterprise focused on understanding the causes of cancer and finding treatments for it. That war has not

DEPRESSION

A *Nature* special issue
nature.com/depression



been won — but no world leader ever stood up and declared a war on depression, and that fact is reflected in the more generous funding that cancer still receives. Garen Staglin, co-founder of One Mind, a non-profit organization in Seattle, Washington, that funds mental-health research, estimates that the US public donates about \$1 billion a year to support cancer research and patients. Mental-health research typically nets less than one-fifth of that.

Campaigning takes energy and confidence — and the very nature of depression makes it difficult for those with the condition to come forward and campaign for support. But another major factor is the long-standing stigma associated with depression. Many people still do not acknowledge that it is a legitimate condition, says Nelson Freimer, a psychiatric geneticist at the University of California, Los Angeles. “A large proportion of people believe depression is just something that we all feel,” he says. “They think you should pull your socks up and get back to work.”

Cancer, too, once carried a stigma. “People didn’t want to talk about their cancer,” says Staglin. “They called it the C-word.” That has changed, he says, as treatments improved, advocacy groups raised awareness and more people spoke out about their battles with the disease. It helped, too, that the reality of cancer is easy to grasp: tumours can be seen, monitored and removed. No such certainty exists in

depression, where the affected tissue is locked inside the brain, cannot be easily seen and certainly cannot be cut out. A rigorous diagnosis requires a two-hour session with a psychiatrist, and yet two patients diagnosed with major depressive disorder — which is how psychiatrists label depression — can exhibit completely different symptoms. “Even one person can have two depressive episodes and the second time

“It’s hard for crackpots to say that pancreatic cancer or breast cancer is not real. Yet somehow they can say that people with mental illness don’t have a real illness.”

is unrecognizable from the first,” says Tim Dalgleish, a clinical psychologist at the MRC Cognition and Brain Sciences Unit in Cambridge, UK. All this leaves the concept of depression as a disorder vulnerable to attack. “It’s hard for crackpots to say that pancreatic cancer or breast cancer is not real,” says Eric Nestler, a psychiatrist and neuroscientist at the Icahn School of Medicine at Mount Sinai in New York City. “Yet somehow they can say that people with mental illness don’t have a real illness. It really is awful.”

Efforts are under way to change how depression is defined and diagnosed in research. Last year, Thomas Insel, head of the National Institute of Mental Health in Bethesda, Maryland,

pushed researchers funded by the institute to eschew classical psychiatric diagnoses, which tend to be indistinct and overlap. Instead, a study might group together patients with specific symptoms, such as anxiety or difficulty with social communication, that are linked to depression as well as to other psychiatric disorders. The hope is that focusing on well-defined traits will reduce some of the experimental noise from artificial diagnostic boundaries, eventually leading to new diagnoses that are grounded in biology. “Ultimately, depression is as biological as cancer and heart disease; it is simply a matter of identifying the relevant molecules,” says Nestler. “It just turned out to be a lot harder than any of us thought it would be decades ago.”

GENETIC PROMISE

Some researchers hope that genetics will help to define depression and delineate subgroups within the condition. That has been the case in cancer, where in the past few years many countries have poured money into the analysis of genomes from a wide range of cancers. The results are revolutionizing the field: they have generated a huge list of mutations linked to cancers, some of which can now be used to match a patient to a therapy. It is a revolution still in progress, but it has placed cancer at the leading edge of personalized medicine.

Depression studies have not fared as well. The largest study so far — a search through

the genomes of just over 16,000 patients with major depressive disorder and another 60,000 controls — has turned up just one, as yet unconfirmed, genetic association¹. Jonathan Flint, a psychiatrist at the University of Oxford, UK, who has been looking for genetic links to depression for nearly two decades, says that some colleagues ask him why he is still working on the problem. “What has held back the entire field is the belief that it’s intractable,” he says. “What is the point of doing something if you’re not going to get anywhere with it?”

The problem stems — yet again — from the disorder’s fuzzy definition: grouping everyone with a diagnosis of major depressive disorder into one genetic study is like looking for the genetic risk factors for fever, explains Flint. “You would have lumped together autoimmune disease, infection, cancer and a whole set of different conditions.” And it is not clear that more funding a few decades ago would have helped the field to move much faster, he says, because the genomic technologies needed for such studies have become available only in the past ten years. But even since then, cancer studies have far outstripped those for depression. “Surely we can do better,” he says. “We have to do better.”

Scientists are already doing better in identifying the genes that underlie some other mental-health disorders, such as schizophrenia. Like depression, schizophrenia can be difficult to diagnose accurately, and initial attempts to find genetic risk factors yielded few hits. But an international group of researchers known as the Psychiatric Genomics Consortium worked to ramp up the sample size in the hope of increasing statistical power and helping the signals to rise above the noise. In September, the consortium published an analysis² of nearly 40,000 genomes from people with schizophrenia that together highlighted 108 different regions potentially linked to the disorder. The consortium now plans to do the same for depression, aiming to scrutinize up to 60,000 genomes from people with the condition.

ANIMALS ON TRIAL

Results from genetic studies could help depression researchers to clear another major hurdle: the development of better animal models. Scientists studying cancer now have a rich choice of model animals that form a crucial part of their research. These include mice that have been engineered to express cancer-associated genes found in human tumours, and even ‘personalized’ animal models that have been tailored to study a person’s disease by transplanting a piece of their tumour into the mouse. Depression researchers, however, have faced huge challenges in creating mice or other animals that behave in a way that mirrors how people are affected by the disorder (see page 200).

Those who do study depression in animals often use physical stresses to prompt behaviours

seen in people with depression. The most common assay is the ‘forced swim test’, in which mice are plunged into water and timed to see how long they struggle to get out. (Those that give up sooner are taken to have depression-like behaviour.) The assay has been used to screen drug candidates — and many antidepressants on the market do extend the time that a mouse is willing to fight. But it is far from ideal: human depressive episodes are rarely triggered by physical stress, and there are signs that antidepressants act differently in this model compared

“What has held back the entire field is the belief that it’s intractable.”

In an attempt to mimic what happens in humans more closely, Nestler and his colleagues subject mice to chronic social — rather than physical — stress. In this ‘social defeat’ model, the researchers place a mouse in a cage with a “bigger, meaner mouse”, he says. The bigger mouse starts to beat up the smaller one, and the fighting continues until the researchers separate the mice using a screen. After ten days of fighting, the smaller mouse typically no longer shows interest in pleasurable activities such as sex or drinking sugar water, and avoids social contact, even with litter-mates³. This reflects some of the symptoms shown by people with depression. So far, the social defeat model seems to better mimic the action of antidepressants in humans, says neuroscientist Ming-Hu Han, also at the Icahn School of Medicine. Experimental drugs that act quickly in people, for example, also work rapidly to ease responses to social defeat in mice.

Mental-health researchers acknowledge that even the best animal models remain a crude reflection of a complicated human disorder. “To understand human circuitry, it isn’t just about whether you will seek out sugar water,” says Helen Mayberg, a neurologist at Emory University in Atlanta, Georgia. “There’s guilt, there’s suicide.” It is also difficult to use animals to study the placebo effect, which is particularly prominent in depression studies and complicates clinical trials of potential antidepressants.

Some scientists question whether an animal can ever truly mimic the human condition. “I don’t like to say I study depression because I don’t think that can be done in animals,” says Olivier Berton at the University of Pennsylvania in Philadelphia. “Those representations of disease are hurting the field and we need to forget them.” Instead, Berton says that he studies stress responses in mice.

There is one way in which the science of cancer and depression are closely aligned, and that is in the growing appreciation of their complexity. Genetic studies of tumours are showing that they are not just divided into lung, liver and other tissue types, but that each tumour is an

intricate mosaic of cells with different mutations and behaviours, and that this mosaic differs from one person to another (see *Nature* **464**, 972–974; 2010).

In depression, an equally complicated picture is beginning to emerge. Researchers always knew that understanding it would be difficult — this is the brain, after all. But as they sort through the thousands of different kinds of neuron in the brain, it is becoming clear that it is important not only to identify the cells, but also to find how they are connected to one another in circuits. Efforts now under way to understand neural circuits may not have happened any earlier, even if depression research had been funded to cancer levels, says Nestler. Picking them apart requires methods that did not exist until recently — for studying single cells, mapping neural connections and activating specific brain circuits.

“We lacked some of the basic knowledge and tools of the brain,” he says.

CIRCUIT TESTING

Now, with those tools in hand, researchers are deep into dissecting the neural circuits involved in depression and working out how to manipulate them using methods that rely on magnets or electrical current. Such work could point to treatments that go beyond the traditional antidepressant pill, says Noah Philip, a psychiatrist at Brown University in Providence, Rhode Island. “Treating depression isn’t as simple as filling up a tank of neurotransmitters,” he says. “It’s correcting a disorder of different neural networks that are not behaving properly.” Mayberg’s team, for example, has been testing deep-brain stimulation as a means to relieve depression. Initial studies found a response rate of around 75%, she says, and she hopes to raise that rate using new imaging techniques to guide the surgery.

Nestler and other researchers argue that it would have been premature to declare a war on depression in the 1970s — but that now, with techniques coming online for brain research, could be the right time. “This is still going to take a couple of decades,” he says. “But I have complete confidence that it will work.”

One of the biggest challenges for the field is to spread that confidence and attract more bright scientists to tackle depression, however thorny the problem may seem. “You don’t throw your hands up because it’s intractable,” says Kelsey Martin, a neuroscientist at the University of California, Los Angeles. “You figure out the best way to find a route into the problem.” ■

Heidi Ledford writes for *Nature* from Cambridge, Massachusetts.

1. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium *et al.* *Mol. Psychiatry* **18**, 497–511 (2013).
2. Schizophrenia Working Group of the Psychiatric Genomics Consortium. *Nature* **511**, 421–427 (2014).
3. Berton, O. *et al.* *Science* **311**, 864–868 (2006).



A CHANGE OF MIND

Cognitive behavioural therapy is the best-studied form of psychotherapy. But researchers are still struggling to understand why it works.

Anna's life began to unravel in 2005 when her husband of 30 years announced that he had fallen in love with another woman. "It had never even occurred to me that my marriage could ever end," recalls Anna, a retired lawyer then living in Philadelphia, Pennsylvania. "It was pretty shocking."

Over the course of several months, Anna stopped wanting to get up in the morning. She felt tired all the time, and consumed by negative thoughts. "I'm worthless. I messed up everything. It's all my fault." She needed help, but her first therapist bored her and antidepressants only made her more tired. Then she found Cory Newman, director of the Center for Cognitive Therapy at the University of Pennsylvania, who started her on a different kind of therapy. Anna learned how to obsess less over her setbacks and give herself more credit for her triumphs. "It was so helpful to talk to someone who steered me to more positive ways of thinking," says Anna, whose name has been changed at her request.

BY EMILY ANTHES

Cognitive therapy, commonly known as cognitive behavioural therapy (CBT), aims to help people to identify and change negative, self-destructive thought patterns. And although it does not work for everyone with depression, data have been accumulating in its favour. "CBT is one of the clear success stories in psychotherapy," says Stefan Hofmann, a psychologist at Boston University in Massachusetts.

Antidepressant drugs are usually the first-line treatment for depression. They are seen as a quick, inexpensive fix — but clinical trials reveal that only 22–40% of patients emerge from depression with drugs alone. Although there are various approaches to psychotherapy, CBT is the most widely studied; a meta-analysis¹

published this year revealed that, depending on how scientists measure outcomes, between 42% and 66% of patients no longer meet the criteria for depression after therapy.

But no one knows exactly how CBT helps. Depression is a complex disorder that manifests in many different ways, and CBT is multifaceted, involving a series of talking sessions whose precise content differs from one therapist and patient to another. Working out exactly how it affects the brain requires studies that are difficult to conduct and to fund. Still, researchers are beginning to piece together answers using a combination of clinical psychology and neuroimaging experiments. Learning more about how CBT works — and why it does not work for everyone — could ultimately help doctors to deliver better care.

"If we don't understand the active ingredients, it's going to be hard to improve the treatment," says Daniel Strunk, a psychologist at Ohio State University in Columbus. "In understanding the mechanism, we might ►

DEPRESSION

A *Nature* special issue
nature.com/depression

GYN9037/SHUTTERSTOCK

► understand that that's more relevant to some patients than others."

CBT encompasses a range of psychotherapies, all based on the premise that people with depression have excessively negative, and often inaccurate, beliefs about themselves and the world. It is designed to equip patients with the skills they need to "become their own therapists", says Strunk, by critically examining those negative beliefs. Correct a person's way of thinking, and the depression will lift, according to the theory.

On the face of it, this seems to be supported by results. "There are dozens of studies that show that people after being treated with cognitive therapy — and after being less depressed — show less negative thinking," says Robert DeRubeis, a psychologist at the University of Pennsylvania. "That's as easy to demonstrate as it is that gravity exists."

What scientists debate is the mechanism. People treated with antidepressants and other kinds of psychotherapy also show more positive thinking after recovering from depression. So does changing someone's thought patterns actually cause their depression to lift? Or does the therapy relieve depression in a different way — by helping people to form a bond with a therapist, for example — so that the positive thinking is merely a consequence of their improved mental health?

UNRAVELLING MECHANISMS

To get at this question, researchers have attempted to show that the change in thinking precedes and predicts the gains in mental health. "By the end of treatment, lots of things will be better," Strunk says. "And so what you really want to do is get inside the moment or moments when someone has a positive therapeutic change and try to understand what's shifting in just those moments."

Research by DeRubeis and his colleagues has revealed^{2,3} that many depressed adults undergoing CBT experienced 'sudden gains', in which their symptoms lessened significantly between two therapeutic sessions. These rapid changes accounted for more than half of the patients' total improvement over the course of treatment.

Recordings of the therapy session just before a sudden gain reveal that a person's way of thinking changes more in that session than in others. "Patients are beginning to speak about changing their minds about many of the negative, exaggerated ways in which they've seen their lives," says DeRubeis. The fact that they begin demonstrating these cognitive changes just before their symptoms improve suggests that altering a person's thinking style may indeed lead to recovery.

Researchers have also shown that learning mental coping skills may be the most important kind of cognitive change during CBT. "The cardinal skill is catching your thoughts in a moment where your mood takes a turn for the worse and thinking through the accuracy

"COGNITIVE THERAPY TEACHES YOU TO STEP IN AND USE YOUR PREFRONTAL CORTEX RATHER THAN LETTING YOUR EMOTIONS RUN AWAY WITH YOU."

of your thoughts in that moment," says Strunk.

In CBT, therapists will often ask patients to monitor their own thoughts. Anna, for example, had been doing volunteer teaching when she started therapy, and she often found herself comparing her abilities with those of another teacher, which made her feel inadequate. Anna's therapist asked her to describe what she and the other teacher did during classes. "It was perfectly obvious once you started thinking about it that the other teacher and I both had good moments and bad moments," Anna recalls. "I had begun to focus on every possible negative about what I did, and every possible positive about someone else."

Anna is no longer in therapy, but when such thoughts return, she can now identify them and examine them to determine whether she is drawing unrealistic conclusions. "It's not that I never have sweeping negative thoughts," she says, "but I'm not so much a victim of them."

Strunk and his colleagues have found⁴ that gaining new cognitive coping skills correlates with an improvement in depressive symptoms, even if negative beliefs remain. And once people have learned these skills, they can use them for the rest of their lives, which may explain why the benefits of CBT last even after the treatment has ended — something that has not proved true for antidepressants. There are not many data on whether other forms of psychotherapy have a similarly protective effect.

Some researchers have looked to neuroimaging to understand more about the mechanisms of CBT. People with depression tend to have detectable differences in two primary brain systems: the prefrontal cortex, which is responsible for complex mental tasks such as self-control and planning, and the limbic system — including the amygdala — which is involved in emotional processing. In healthy people, the prefrontal cortex can inhibit amygdala activity, keeping emotions in check. But imaging shows that in many people with depression, the prefrontal cortex seems to be less active. "Depressed people have what you might think of as a trigger-happy amygdala,"

says Greg Siegle, a neuroscientist at the University of Pittsburgh in Pennsylvania.

There is some evidence that CBT can correct these problems. In a study⁵ published in 2007, Siegle used functional magnetic resonance imaging (fMRI) to show that depressed adults had increased levels of activity in the amygdala when performing an emotional task and reduced levels of activity in the dorsolateral prefrontal cortex when performing a cognitive task. When he followed up 9 of the trial participants, Siegle found⁶ that a 14-week course of cognitive therapy had almost completely reversed the situation.

"The imaging data have been really encouraging," says Timothy Strauman, a psychologist at Duke University in Durham, North Carolina, who has generated similar results⁷. "We do find evidence for the kinds of changes that you would expect."

Researchers speculate that CBT — with its focus on controlling thoughts — re-engages the underactive prefrontal cortex, which, in turn, helps to quieten the hyperactive limbic system (see 'Desired behaviour'). "Cognitive therapy teaches you to step in and use your prefrontal cortex rather than letting your emotions run away with you," says Siegle.

Still, there are caveats. The prefrontal cortex and the amygdala are not the only brain areas that behave differently in depression; nor are they the only areas that therapy affects. The studies have been small and they occasionally contradict each other. "I'm always cautious, because there haven't been enough studies," says Cynthia Fu, a neuroscientist at the University of East London, UK. She estimates that there have been three to four times as many neuroimaging studies of antidepressant use as of psychotherapy. "It's a new field, and people use different tasks, they scan people at different times, and that can lead to quite different results."

TOUGH QUESTIONS, TOUGH ANSWERS

As with the changes in negative thinking, it is not yet clear whether these neurological changes are a cause or a consequence of recovery. To answer that, scientists will need to image patients repeatedly throughout the course of CBT to track the changes and determine whether they predict improvement.

Such studies are expensive, time-consuming and burdensome for patients. And in general, scientists say, it can be harder to study CBT than antidepressants: there are more confounding factors, because therapists and sessions can vary widely. And it is difficult to administer a placebo. Researchers can compare participants receiving CBT with others who are randomly assigned to receive drugs or pharmacological placebos, or are put on a waiting list for treatment; they can also use a control group made up of patients who receive general counselling, but it is impossible to blind the study properly, because the therapists will know which participants are receiving which treatments. Another

challenge is finding funding. “It’s a David-and-Goliath-type comparison,” says Strunk. “There are pharmaceutical companies that make antidepressants, so there’s a lot of financial backing there. The amount of money is just different scales.”

Scientists are eager to know why CBT works for only some people, and to be able to identify those who will respond. “There’s a lot of trial and error,” says Lena Quilty, a psychologist at the Centre for Addiction and Mental Health in Toronto, Canada. But every failed treatment can extend a person’s suffering and add to the cost of medical care.

Certain clinical and demographic factors

seem to predict whether a person will respond to drugs or CBT. Those with personality disorders in addition to depression, for instance, tend to do better with antidepressant medications than with cognitive therapy, and married people seem to benefit more from cognitive therapy than from medication.

Researchers have now begun searching for patterns of brain activity that can identify how well someone will respond to CBT. In a study⁸ published last year, neurologist Helen Mayberg of Emory University in Atlanta, Georgia, and her colleagues used positron emission tomography (PET) to measure glucose metabolism in the brains of 82 adults with depression. They

then randomly assigned each participant to receive 12 weeks of treatment with either CBT or a commonly prescribed antidepressant from the selective serotonin reuptake inhibitor class. People with high activity in the right anterior insula, a brain region that communicates with both the amygdala and the prefrontal cortex, tended to respond well to the drug. Those with an underactive insula were more likely to improve with CBT.

It is not yet clear why. “Fundamentally, one has to think about the network in the brain that goes wrong in depression as being dynamic,” says Mayberg. “Those systems are broken differently in people who respond to different treatments.” Cognitive therapy may be able to repair some problems, whereas drugs may be better at patching up others.

Neuroimaging is not yet a practical tool for identifying the best course of treatment, so researchers are working on other approaches. In 2011, Siegle and his colleagues demonstrated⁹ that clinicians might be able to use the pupil of the eye as a window on the brain. Depressed adults were shown negative words; those whose pupils did not dilate much had reduced activity in some areas of the prefrontal cortex and were more likely to benefit from cognitive therapy than those whose pupils dilated more.

In practice, clinicians hoping to tailor treatments may need to consider many factors, including a person’s marital status, brain activity and genetics; a small body of research indicates that people with certain genetic sequences are more likely than others to respond to CBT.

And scientists may need to get a better handle on depression itself — in all its forms and manifestations — before they can understand how CBT alleviates it and who is likely to benefit. Strauman is optimistic about the growing number of collaborations that he is seeing between neuroscientists and clinical psychologists who are willing to tackle the problem. “I think we’re finally at the point,” he says, “where the complexity of our thinking is a match for the complexity of the disorder.” ■

Emily Anthes is a science journalist in New York City.

1. Cuijpers, P. et al. *J. Affective Disorders* **159**, 118–126 (2014).
2. Tang, T. Z. & DeRubeis, R. J. *J. Consult. Clin. Psychol.* **67**, 894–904 (1999).
3. Tang, T. Z., DeRubeis, R. J., Beberman, R. & Pham, T. *J. Consult. Clin. Psychol.* **73**, 168–172 (2005).
4. Adler, A. D., Strunk, D. R. & Fazio, R. H. *Behav. Ther.* <http://dx.doi.org/10.1016/j.beth.2014.09.001> (2014).
5. Siegle, G. J., Thompson, W., Carter, C. S., Steinhauer, S. R. & Thase, M. E. *Biol. Psychiatry* **61**, 198–209 (2007).
6. DeRubeis, R. J., Siegle, G. J. & Hollon, S. D. *Nature Rev. Neurosci.* **9**, 788–796 (2008).
7. Ritchey, M., Dolcos, F., Eddington, K. M., Strauman, T. J. & Cabeza, R. *J. Psychiatr. Res.* **45**, 577–587 (2011).
8. McGrath, C. L. et al. *JAMA Psychiatry* **70**, 821–829 (2013).
9. Siegle, G. J., Steinhauer, S. R., Friedman, E. S., Thompson, W. S. & Thase, M. E. *Biol. Psychiatry* **69**, 726–733 (2011).

DESIRED BEHAVIOUR

Depressed individuals often have an overactive (pink) amygdala and a less-active (blue) prefrontal cortex compared to healthy people.

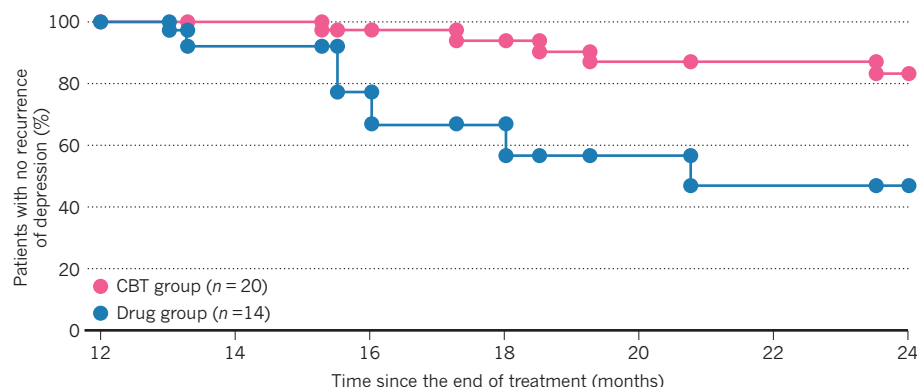
1 Activity in the amygdala, which is associated with emotional responses, may repress activity in the prefrontal cortex, which is involved in executive function.

2 Some researchers suspect that antidepressant medication quiets activity in the amygdala, whereas cognitive behavioural therapy (CBT) increases activity in the prefrontal cortex.

3 Although the two approaches work in different ways, they both seem to restore stability in these areas of the brain.

LONG-TERM BENEFITS

Studies have shown that people who completed a course of CBT had a lower rate of relapse than people who stopped taking antidepressant medication.



COMMENT

HEALTH Britain backs clinics for isolating suspected Ebola cases in Sierra Leone **p.192**



FILM Alan Turing and Stephen Hawking biopics compared **p.195**

SCI-FI Physicist Kip Thorne on his Hollywood experience with *Interstellar* **p.196**

FUNDING Grant applicants deserve to see full expert feedback **p.198**

ILLUSTRATION BY DENIS CARRIER



Depression needs large human-genetics studies

To understand the molecular mechanisms of depression, collect genetic data from more than 100,000 people, says **Steven Hyman**.

Five decades of work on antidepressant drugs have not made them more likely to lift people out of depression. Medications and psychotherapies help some people with the disease, but fail many others. In a study¹ conducted in real clinical settings, a common antidepressant produced a response in just under

half of the participants, and achieved full remission in only 28%.

Failures to improve efficacy reflect

DEPRESSION

A *Nature* special issue
nature.com/depression

continued ignorance of the molecular mechanisms of depression. Progress in other disorders, such as autism and schizophrenia, suggest that the best hope for insights is to identify specific genetic variants associated with the disease. However, success in depression will require studies of much larger collections of ►

► human DNA samples than in other diseases if statistically significant signals are to come through. The investments will be well worth it.

SERENDIPITY'S CURSE

Today's antidepressants all descend from serendipitous observations in human trials of two molecules more than 50 years ago: imipramine, synthesized as a potential antipsychotic drug, and iproniazid, designed to treat tuberculosis. Neither worked for its intended purpose, but both relieved the symptoms of depression. More than 20 antidepressants followed, all working in roughly the same way as these two — boosting signalling by neurotransmitters, mostly noradrenaline and serotonin. Although newer drugs are less toxic, none has greater efficacy than the first drugs of the 1950s².

What went wrong? Researchers attempting to make new antidepressants developed rodent behavioural assays based on the effects of imipramine and iproniazid on healthy laboratory animals. Unsurprisingly, these assays selected new compounds that act similarly to the drugs on which they were based².

To find more-effective drugs that help more people, we must look for drugs that work in different ways. In the past decade, some investigators have returned to observing drug effects in depressed people and found, for instance, that a dose of intravenous ketamine (which blocks one of the receptors for the neurotransmitter glutamate) relieves symptoms of depression rapidly and for up to a week. Although potentially leading to new treatments, observations of this sort are unlikely to yield deep molecular insights into the mechanisms of depression.

Any drug that perturbs several circuits and cell types in a complex system such as the brain yields an enormous number of divergent effects. Which of these myriad effects matters? Without independent clues of what happens in the brain to produce depression, researchers cannot seek out those drug effects that might counteract the pathology. After 50 years, we still do not know very much about how imipramine (sometimes) reverses the symptoms of depression.

HURDLES TO OVERCOME

Depression is difficult to study: it is hard to find features of the brain to focus on. The plaques and tangles observed in the brains of people with Alzheimer's disease provided important early clues to amyloid- β and tau protein. In contrast to many other brain disorders, depression lacks distinctive pathology, not even the thinning of the brain's cortex seen in schizophrenia.

For now, animal models are more likely

to be a source of distraction than insight. The cellular and molecular mechanisms — and thus possible drug targets — by which humans exert cognitive control over thought, emotion and behaviour differ markedly from those in mice and rats. Rodents lack the brain region crucial for

“The tiny signal-to-noise ratio has doomed past searches for candidate genes.”

human cognitive control, the dorsolateral prefrontal cortex.

Deciphering signals in humans will be tough. People with major depressive disorder (MDD), according to the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, span diverse ages of onset, mixtures of symptoms and impairments, and show varied treatment responsiveness. A person can be diagnosed with major depression by meeting any five of nine criteria. Thus, some participants enrolled in a study might overlap on only a single, varying criterion. Moreover, in the *DSM-5*, the threshold for being considered ill was determined arbitrarily. Unlike hypertension or blood pressure, there are no empirical studies of outcome that set the bar for a medical diagnosis.

HIDDEN CLUES

Amid these traps, one clue to molecular mechanisms of depression has long beckoned: genetic analysis. Spurred by recent technological developments, genetic analysis has begun to yield powerful clues in autism, schizophrenia and bipolar disorder^{3,4}.

But unravelling the genetics of depression also faces significant obstacles. The heritability — the fraction of phenotypic variance explained by genes — is 35% for MDD, compared with 65–80% for the psychiatric diseases for which genetic analysis is proving very successful^{3,4}. Several lines of evidence argue that risk of depression is highly polygenic, involving many hundreds of genes. This, combined with diagnostic ‘noise’ — that is, the heterogeneity and misdiagnosis of people with MDD — drives up the number of subjects needed to find significant genetic associations. Combined with our ignorance of disease mechanisms for depression, the tiny signal-to-noise ratio has doomed past searches for candidate genes and gene–environment interactions⁵.

The largest meta-analysis of genome-wide association studies of depression (approximately 9,500 cases) has yielded no significant findings⁶. Similarly sized studies of almost all other conditions have convincingly implicated⁷ at least some genetic loci (see ‘Signal search’). So far, 108 independent loci have been found to demonstrate genome-wide significance for schizophrenia⁴.

Nonetheless, I am convinced that genetic variants for depression can be found. Type 2 diabetes also has a heritability of 35% or lower. Perhaps because this type of diabetes can be diagnosed simply and objectively, studies with fewer than 10,000 cases have identified many common genetic variants. More than 100,000 people with MDD will be needed to find enough loci to inform biology and therapeutics. Amassing a data set of this scale is difficult, but worthwhile and possible. A meta-analysis⁸ of genome-wide association studies published earlier this year for adult height included more than 250,000 subjects and has found 697 common variants thus far, explaining nearly one-fifth of the heritability.

SCIENTIFIC TOEHOLDS

Diagnosing MDD is more challenging than measuring height, of course, but the extra effort and cost are justified to illuminate the world's greatest source of disability. We will need a global coalition to assemble the requisite number of samples. An assessment of common genetic variants for psychiatric disorders now costs less than US\$100 per person; recruiting participants and the logistics of obtaining and shipping DNA samples have costs, but samples can be made available from many existing studies of depression and from registries in Scandinavia. Samples of other ethnicities are also being assembled.

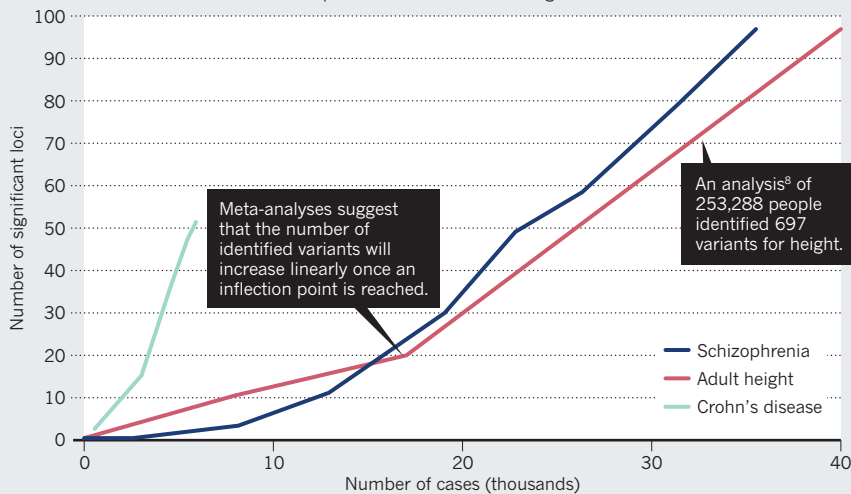
It took about five years for the Psychiatric Genomics Consortium to collect and analyse nearly 40,000 schizophrenia cases. Similar efforts in depression are gearing up, and should proceed smoothly because the steps in the genetic analysis of psychiatric disorders are now well understood, a microarray (gene chip) fit for the task is available, and experienced scientists are prepared to perform the analysis.

Once genetic loci are identified and shared with the scientific community, many investigators, drawn by the emergence of biologically actionable clues, will rush to put the information to work. This is what has happened in the past decade in autism research. When I was director of the US National Institute of Mental Health, from 1996 to 2001, parent groups and the US Congress strongly encouraged me to increase investments in autism research. Seeing no clear scientific paths forward I was reticent, and requests for proposals generally yielded weak grant applications.

However, genetic insights⁹ transformed a featureless landscape into one with real scientific toeholds. This work has not yet yielded new treatments, but there is now a path. New genetic insights even led to clinical trials of treatments for Fragile X syndrome, which has both autism and intellectual disability as cardinal features.

SIGNAL SEARCH

Studies involving thousands of people can be required to find the genetic variants, or loci, associated with diseases such as schizophrenia or traits such as height.



Although these trials were unsuccessful, drug companies have now begun to consider this condition as amenable to pharmacological treatment.

Critics will counter that the genetic discoveries among certain autism spectrum disorders are rare mutations of large effect, and those to be expected in depression will be common variants with small effects on risk. The answer is that definitively identified risk alleles will point to genes involved in pathogenesis and these genes will, in turn, identify biological pathways that provide potential targets for therapeutics. Risk alleles also show whether candidate drugs should increase, decrease or otherwise alter the activity of the implicated pathway. In 2012, a variant in the gene that produces β -amyloid precursor protein was identified as protecting against Alzheimer's disease, providing powerful evidence that drugs targeting amyloid peptides might work.

Genetic loci for depression will be especially useful when considered in the context of variants that contribute to other neuropsychiatric disorders. Just as inflammatory diseases and autoimmune diseases have both shared and unshared risk alleles, major depression will have risk variants that overlap with bipolar disorder, schizophrenia and several other psychiatric disorders. Comparing the effects of genes shared across diseases and those associated with just a single disease will hopefully be informative¹⁰.

To assess the biological significance of the genetics, reprogramming human skin fibroblast cells into neurons and

gene-editing technologies will be hugely useful. Human skin cells can now be coaxed into forming increasingly specific types of neuron, and risk variants can be engineered into or removed from cells. As a result, researchers can interrogate hundreds of variants for their effects on human cells and, for the first time, screen drugs for psychiatric disorders *in vitro*. It will be difficult to relate cellular phenotypes to the symptoms of depressed individuals, but new technologies to examine circuits *in vitro*, in animals and in people could help.

Although the individual effects of specific risk alleles will probably be very low, the information they could yield is enormous. They will allow researchers to rely on understanding, and not just luck, to make inroads into one of the biggest causes of human misery. ■

Steven Hyman is director of the Stanley Center for Psychiatric Research at the Broad Institute in Cambridge, Massachusetts, USA. e-mail: seh@harvard.edu

1. Trivedi, M. H. *et al.* *Am. J. Psychiatry* **163**, 28–40 (2006).
2. Hyman, S. E. *Sci. Transl. Med.* **4**, 155cm11 (2012).
3. Sullivan, P. F., Daly, M. J. & O'Donovan, M. *Nature Rev. Genet.* **13**, 537–551 (2012).
4. Schizophrenia Working Group of the Psychiatric Genomics Consortium *Nature* **511**, 421–427 (2014).
5. Duncan, L. E. & Keller, M. C. *Am. J. Psychiatry* **168**, 1041–1049 (2011).
6. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium *Mol. Psychiatry* **18**, 497–511 (2013).
7. Levinson, D. F. *et al.* *Biol. Psychiatry* **76**, 510–512 (2014).
8. Wood, A. R. *et al.* *Nature Genet.* <http://dx.doi.org/10.1038/ng.3097> (2014).
9. Jamain, S. *et al.* *Nature Genet.* **34**, 27–29 (2003).
10. Cross-Disorder Group of the Psychiatric Genomics Consortium *Nature Genet.* **45**, 984–994 (2013).



SYLVAIN CHERKAOU/COSMOS/EYEVINE

A nurse attends to a patient at a Médecins Sans Frontières Ebola treatment centre in Kailahun, Sierra Leone.

Tough choices to reduce Ebola transmission

Christopher J. M. Whitty and colleagues explain why the United Kingdom is funding many small community centres to isolate suspected cases in Sierra Leone.

The Ebola outbreak in West Africa is larger than all previous Ebola outbreaks combined, and is still expanding¹. With a death toll in the thousands, and mortality at around 70%, it has undermined fragile health-care systems by filling hospitals with highly infectious patients and killing health workers.

Transmission in dense urban populations presents challenges never before seen. Vaccines and drugs would offer a complementary approach to control and they must be fast tracked. But these medicines are not available for immediate widespread use. To quell this unprecedented outbreak, we have to use methods that have not been tested.

Experience from past outbreaks has established reliable methods to control transmission in hospitals and at funerals of people who die from Ebola — two of the three main venues for transmission (see 'Hospitals and funerals'). These efforts remain essential, and the scientific and operational strands must complement one another.

What most distinguishes the current situation from previous outbreaks is the high proportion of transmission occurring in the community. No techniques have been developed to control community transmission at this scale^{2,3}. Notwithstanding uncertainties, the global community must act now.

The UK government is leading the international response to Ebola in Sierra Leone, providing technical, financial and logistical help. This article sets out the scientific basis for the UK government's strategy to assist Sierra Leone's government to reduce transmission. In addition to substantially scaling up conventional capacities at hospitals, we plan to help to build and support community isolation centres where people can voluntarily come to be isolated if they suspect that they have the disease.

These interventions represent uncertain ground, but hesitation is more dangerous than trying out potentially ineffective methods (see 'Delays mean more deaths'). Climbing rates of infection could soon overwhelm

Sierra Leone's already challenged health services, and so remove any realistic chance of public-health intervention ending the outbreak. The Sierra Leonean health system and the many brave local and international health staff need help to lower transmission to the point at which mopping up remaining islands of high transmission will be possible.

Until the outbreak is over, we will not know whether we have launched the best response. We invite critiques and suggestions, but must act swiftly. Further delay will result in more infections and deaths, and only sabotage future efforts.

THE CHALLENGE

About 500 new cases of Ebola are being diagnosed every week. This number is expected to double in the next 30 days⁴. The force of transmission of a disease outbreak is quantified as R , the average number of people infected by each newly infected person (the reproduction number). When $R = 1$, an outbreak is neither increasing nor decreasing.

EBOLA TRANSMISSION

Hospitals and funerals

The initial symptoms of Ebola (which may include fever, diarrhoea and vomiting) mimic those of many common diseases, including malaria, pneumonia and gastrointestinal infections. Transmission occurs through bodily fluids: diarrhoea, vomit, blood and probably sweat and semen. Patients become infectious when they are symptomatic, and remain highly infectious until they begin to recover. Corpses are highly infectious.

Controlling transmission requires minimizing contact with bodies, bodily fluids and contaminated items such as mattresses, clothes and clinical waste. Transmission does not occur through breathing or insects. Transmissibility through existing modes of infection could evolve as the outbreak expands.

Ending previous Ebola outbreaks relied on stopping transmission in hospitals through strict infection control; adapting funeral rites to reduce risk of transmission; helping local people to protect themselves; and actively tracking down those likely to be infected⁷.

Infection control in hospitals requires sufficient resources and trained personnel to implement measures, such as using personal protective equipment. In relatively well-equipped hospitals with stringent training and control protocols (such as those run by Médecins Sans Frontières), infection rates remain low.

Reducing transmission from the dead is more complex, and social scientists play an essential part. Protocols for handling and

disposing of a body to minimize the risk of infection are well established, but must be implemented with the cooperation and consent of grieving communities. Funeral rites often require that family members and mourners have extensive contact with the deceased.

Previous Ebola outbreaks in the Democratic Republic of Congo and Uganda have shown that unless infection-control routines can be adapted to local burial practices, communities will continue their traditions. Anthropologists have documented funeral rites in Sierra Leone and other countries. Dialogues with Ebola-stricken communities are essential to adapting their traditions to reduce body handling respectfully.

According to our analyses, R in Sierra Leone is currently between 1.2 and 1.5. In some areas, R is considerably higher. If R remains above 1, any public-health intervention (except a vaccine) will eventually be overwhelmed by the number of new infections. Getting R below 1 is the single strategic aim of the UK effort at this stage of the outbreak.

Reducing R depends on stalling transmission in the community, which depends on reducing the time between when people first show symptoms and are isolated. Delay occurs in two broad settings: patients who have symptoms consistent with Ebola but have not been positively diagnosed; and people diagnosed with Ebola who are being nursed at home because of reticence to move to facilities, or a lack of available health-care beds or transport. Patients in the later stages of Ebola are the most infectious, in part because symptoms such as vomiting and diarrhoea are worse⁵. Models vary, but most work, including that led by N.F. and W.J.E., suggests that isolating 70% or more of Ebola cases within three days of their becoming symptomatic is key to reducing R to 1.

TWO-PRONGED STRATEGY

Designing measures to control community transmission poses epidemiological, clinical and social-science challenges. Currently, most people with early Ebola-like symptoms will not have Ebola, but diseases such as malaria and pneumonia. Undiagnosed individuals should be isolated both from health workers (as in Ebola wards) and from each other, so that the true Ebola cases do not infect those with other diseases. The high number of suspected cases greatly complicates isolation efforts.

Community control of previous Ebola

outbreaks relied on active 'case-finding': tracing contacts of known cases and then monitoring and quickly isolating suspected cases early. It can take as many as ten staff to monitor the contacts of just a single case, and the task is potentially more difficult in mobile urban populations, especially as Ebola case numbers rise.

The ideal approach would be active case-finding combined with isolating patients in fully equipped and staffed hospitals, but this is not practical in the current situation.

"Centres must be able to operate with small, relatively unspecialized staff drawn from the community."

In some areas, the outbreak has already overrun hospitals; many suspected Ebola cases are being turned away for lack of beds. An unknown proportion of cases remain at home. For those who seek medical care, the current median time between becoming symptomatic (and thus infectious) and isolation in Sierra Leone is four days. Many wait more than a week.

These delays will only get longer. Current UK aid efforts — which have so far pledged more than £200 million (US\$320 million) and the largest UK troop deployment outside Afghanistan — will help to increase Sierra Leone's bed capacity threefold by January. But at current measures of R , the projected increase in new cases (thousands per week) will far exceed the number of possible new hospital beds. To avoid that scenario, beginning this month, affected regions must substantially increase rates of early isolation for suspect and confirmed Ebola cases.

One proposed strategy — giving families information and basic personal protective

equipment (PPE) to minimize transmission while nursing patients at home — is problematic. Using PPE safely is difficult even for professionals, as infection rates in health-care workers demonstrate. And identifying cases and training families requires staff that Sierra Leone does not have. This approach is acceptable only as a desperate humanitarian measure when there is no space available in health facilities. It is not a good strategy to reduce transmission.

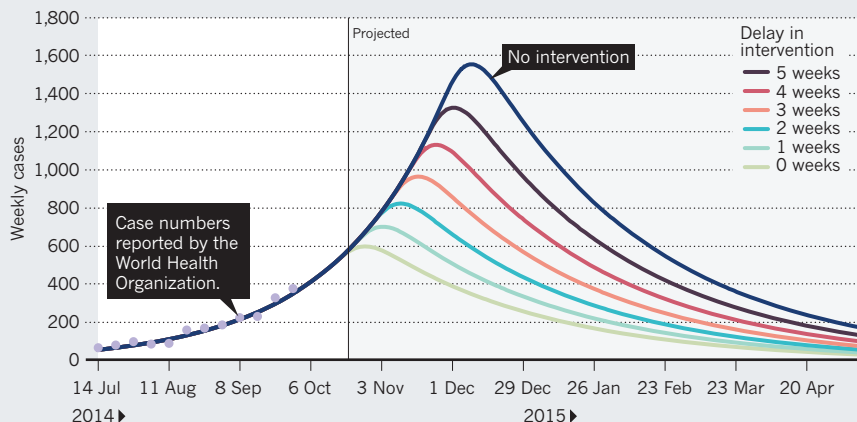
PASSIVE CASE-FINDING

The UK government has decided to support another strategy: passive case-finding with community isolation. This is at the request of the government of Sierra Leone and endorsed by the World Health Organization. The strategy, also called voluntary sequestration, encourages those with suspected Ebola infection to gather in units where they can receive basic health care and avoid infecting members of their families and communities. By quickly isolating suspected cases (and boosting the proportion of such cases isolated early), this approach aims to keep people with Ebola from infecting others, until R is pulled below 1.

Some models are currently being piloted near Sierra Leone's capital, Freetown, co-designed with local clinicians and communities. An initial 200 community units can be expanded according to how and where the outbreak progresses. All share similar requirements. Community-level units must be easy to construct, relying mainly on tents or repurposed buildings in or near affected communities, that can be set up in days or weeks. Many small units are better than a few large ones. They can be placed closer to communities (distance can be a barrier to people

DELAYS MEAN MORE DEATHS

For every week that effective interventions are not implemented in Ebola-stricken areas, the case-number peak will be pushed higher and later, meaning that more hospital beds will be needed.



checking themselves in). It also reduces the risk of a major outbreak if infection control breaks down in one unit. Centres must be able to operate with small, relatively unspecialized staff drawn from the community, supported with basic training and provision of equipment and supplies by agencies such as non-governmental organizations (NGOs).

Most importantly, centres must not become sources of new infections. This means effective isolation, as well as diagnosing and moving confirmed Ebola cases on to specialized treatment facilities as quickly as possible.

Because most patients in the early stages of the disease will have relatively minor symptoms, one health worker can care for several people. And because patients choose to report to these centres, communities are more likely to accept them. Even incomplete coverage could yield successes. The aim is not to find every suspected case, but to attract enough early cases to shorten the time that infectious people spend in the community, and so reduce R .

DATA, MODELLING AND RISK

There is little current evidence from previous outbreaks on how well, or even whether, voluntary community isolation reduces transmission. We will need to learn and adjust as data come in. Accurate, recent, open data are essential to shape an effective response, but data-capture systems were implemented only in September in the worst-affected countries. These now allow us to estimate R more accurately and to track time between symptoms and isolation.

Major uncertainties remain — most notably around the proportion of Ebola cases currently being detected by health systems. If only a small proportion of cases are being identified, then a dramatic increase in self-reporting is essential. Conversely, if the major issue is, as we suspect, delays in isolation, then even a one- or

two-day reduction in that time might be enough to achieve control.

Modelling can help to predict what provisions are needed. Our current projections of the outbreak in Sierra Leone suggest that more than 200 ten-bed isolation units will be required by the end of December. This estimate may be too low.

The UK government has been incorporating social-science advice into the incentives for bringing sick patients to community centres⁶. The desire for good medical care, for protecting loved ones from infection and possibly provision of food or money would all encourage people with early symptoms to go to Ebola evaluation centres. But distance, suspicion, fear and stigma will discourage early reporting. Under-incentivizing would mean that too few people with Ebola will report to centres to reduce transmission. Over-incentivizing could mean that too many non-infected people report, overwhelming facilities and increasing risks of transmission. To achieve the optimal balance, facility planners will need to engage with local populations and be flexible enough to adopt better practices as evidence accumulates.

Careful collaborative work with communities, building on what is already known from anthropological studies and field reconnaissance, must guide how facilities are located, designed, staffed and operated to maximize social acceptance.

This public buy-in is essential. Communities have reacted negatively to the opening of treatment facilities, and in some cases, people have forcibly removed family members from isolation units. Rumours of nefarious killing and body-part theft have circulated, reflecting anxieties that resonate with people's experiences of disease, war and government intervention (see www.ebola-anthropology.net). The murder of health workers and journalists in Guinea this September exemplifies the risks to staff if interventions go wrong.

Optimal design will be learned on the fly. We know how to minimize infection in dedicated Ebola wards with highly trained and supervised staff, but not in facilities with lightly trained staff in which most people do not have Ebola (but some do). Evidence will have to be collected on the best ways to design units, place patients, speed Ebola diagnoses and transport cases. There are likely to be early failures from which the global community must learn. Given the risks that health workers take, getting Ebola vaccines to staff working in the units as soon as they have proved safe is an ethical imperative.

Extrapolating from data collected in neighbouring countries on common pathogens and antimicrobial drug resistance will be essential. Diagnostic testing, except for Ebola, must be kept to a minimum to avoid contamination from blood and sharp instruments. Health workers should use treatments on the basis of symptoms alone to deliver appropriate care for malaria and bacterial disease without testing, a practice that has been used effectively in many under-resourced settings. Faster and safer Ebola tests (those that would not require using blood with the live virus) would substantially improve clinical management.

We hope that in the coming months, safe and effective vaccines will be produced at large enough volumes to stop this outbreak, but thousands are dying now. Qualitative, quantitative and clinical sciences, and the ability to adapt and learn from mistakes, are urgently needed to reduce community transmission. ■

Christopher J. M. Whitty is chief scientific adviser at the UK Department for International Development. **Jeremy Farrar** is director of the Wellcome Trust in London, UK. **Neil Ferguson** is professor of mathematical biology at Imperial College London, UK. **W. John Edmunds** is professor of infectious-disease modelling at the London School of Hygiene & Tropical Medicine, UK. **Peter Piot** is director of the London School of Hygiene & Tropical Medicine, UK. **Melissa Leach** is director of the Institute for Development Studies in Brighton, UK. **Sally C. Davies** is chief medical officer and chief scientific adviser at the UK Department of Health. e-mail: c-whitty@dfid.gov.uk

1. WHO Ebola Response Team *N. Engl. J. Med.* **371**, 1481–1495 (2014).
2. World Health Organization Ebola Response Roadmap Situation Report: 31 October 2014 (WHO, 2014).
3. Kerstiens, B. & Matthys, F. *J. Infect. Dis.* **179** (Suppl. 1), S263–S267 (1999).
4. Borchert, M. et al. *BMC Infect. Dis.* **11**, 357 (2011).
5. Bausch, D. G. et al. *J. Infect. Dis.* **196** (Suppl. 2), S142–S147 (2007).
6. Hewlett, B. S., Epelboin, A., Hewlett, B. L. & Formenty, P. *Bull. Soc. Pathol. Exot.* **98**, 230–236 (2005).
7. Heymann, D. L. *Nature* **514**, 299–300 (2014).



Left, Benedict Cumberbatch as Alan Turing in *The Imitation Game*; right, Eddie Redmayne as Stephen Hawking in *The Theory of Everything*.

FILM

Enigma variations

Robert P. Crease ponders a brace of biopics on Alan Turing and Stephen Hawking.

Early in *The Imitation Game*, a constable says, “I think Alan Turing’s hiding something.” The implication is that it is his sexuality. But as well as being gay, Turing was a master code-breaker, spy and recluse. Secrets drive this thrilling, fast-paced film — as gripping as it is tragic — about one of history’s most brilliant mathematicians and computer engineers. Meanwhile, a movie about physicist Stephen Hawking turns his life into a feel-good saga.

Morten Tyldum’s *The Imitation Game* makes it clear that Turing’s passion was finding solutions to abstract logical problems. But how to animate a character (played by Benedict Cumberbatch) whose life is all in his head? The answer is provided by the Enigma machine, used by Nazi Germany to encipher military messages. In trying to crack the secret, Turing’s otherworldly interests in ‘impossible’ logical challenges intersected with urgent worldly ones — the challenges of the British government, at one of the darkest junctures of the Second World War.

An early scene, in which we are first shown Enigma, dramatizes that intersection. Turing pronounces the device beautiful; a government official declares it “the crooked hand

The Imitation Game

DIRECTED BY: MORTEN TYLDUM
Black Bear Pictures: 2014.

The Theory of Everything

DIRECTED BY: JAMES MARSH
Working Title Films: 2014.

of the devil”. As Turing and his colleagues at Bletchley Park, the British code-breaking centre, work on deciphering Enigma, every tick of the clock means more warfare, more death. The film cross-cuts scenes of bombs raining on cities with scenes of prospective code-breakers solving crossword puzzles, and you get the connection.

When the Bletchley gang finally solve the code, they have to hide their triumph or risk tipping the enemy off. Turing is asked to determine which secret — such as news of impending attacks on convoys or cities — can be acted on without arousing suspicion. Meanwhile, the Soviets are spying on Bletchley, kept at bay with a drip-feed of minor information. The UK government blackmails Turing (gay in an era when that was illegal) to force him to spy on the Soviet ring.

The film takes its title from the Turing test, an imitation game that he devised in which a human has to judge whether a concealed interlocutor is a machine or a human. The idea is invoked in a scene, played powerfully by Cumberbatch, in which Turing is about to reveal the full details of his life story, including the code-breaking and homosexuality. He challenges the (baffled) constable to judge him: “What am I? Am I a criminal? Am I a war hero?”

The movie neither sensationalizes nor ignores Turing’s sexuality. It tells us little about his work in maths or cryptography, but it allows us to glimpse something deeper: how the ordinary world is shaped by mathematicians and scientists in ways that many people hardly suspect. A final scene shows ecstatic celebrations in the street at war’s end. We are left feeling that, if such events fail to give the Turings of the world their due, the revelry is hollow. Turing, of course, was not just unheralded: like most Bletchley employees, his accomplishments were classified and his war records expunged. Later, he was arrested for ‘gross indecency’ and forced to undergo chemical castration. He killed himself in 1954.

► **NATURE.COM**
For Nature’s Alan Turing special, see: nature.com/turing

► *The Theory of Everything* is, like *The Imitation Game*, about a scientist. There the similarities end. The film tells us that Hawking has refused a knighthood, but it thoroughly canonizes him. There have been other depictions of Hawking's life, including *Hawking* (2013), a documentary by Stephen Finnigan, and *Hawking* (2004), a BBC television film starring Cumberbatch. This turn, with Eddie Redmayne in the lead role and directed by James Marsh, is loosely organized around Hawking's relationship with his first wife, Jane Wilde, and based on her memoir *Travelling to Infinity* (Alma, 2008). Despite Redmayne's fine performance, it unfolds with such plodding, reverential linearity, and with so much melodrama and uplift, that it is the least compelling of the treatments.

The story begins in Cambridge, UK, in 1963. Stephen and Jane flirt. He is diagnosed with motor neuron disease and told that he has two years to live. She remains selflessly devoted and they marry. Stephen works despite his disability. "He has done it!" we are told at one point, but we haven't a clue as to what 'it' is. Caring for Stephen and their three children gives Jane moments of doubt and insecurity. The ins and outs of the couple's relationships with each other and with their eventual second spouses — choir director Jonathan Hellyer Jones and Stephen's nurse Elaine Mason — are rolled out with surprisingly little conflict. Everyone is presented as well intentioned and concerned only for Stephen's welfare.

Key moments in Hawking's life are handled simplistically. He announces, "I will write a book," and knocks off *A Brief History of Time* (Bantam, 1988) without strain or editorial conflict; it is an instant best-seller. A final scene shows Hawking, now a venerated celebrity, delivering a talk to an entranced audience. One questioner asks how it is possible for him to live without believing in God. The question seems pointless: in the world of this film, so full of saints and sacrifice, who needs God? But cloaked in the ethereal strains of his electronically synthesized voice, Hawking's response — "Where there is life, there is hope" — brings the audience cheering to its feet. The scene invites us to regard Hawking himself as a miracle.

Dramatizing the lives of scientists, so much of whose work is mentation, is demanding. These two films demonstrate the promise and the perils. ■

Robert P. Crease is professor of philosophy at Stony Brook University in New York, and co-author, with Alfred Scharff Goldhaber, of *The Quantum Moment*.
e-mail: robert.crease@stonybrook.edu



Matthew McConaughey, Anne Hathaway and David Gyasi in *Interstellar*.

Q&A Kip Thorne

Space-time visionary

Thanks to theoretical physicist Kip Thorne, real science is embedded in Christopher Nolan's film *Interstellar*, in which explorers seek a new home for humankind. Thorne talks about what he learned from the film's unprecedented visualizations of black holes and wormholes, what it and his accompanying book can teach, and the likelihood of humans escaping the Solar System.



How did *Interstellar* come about?

I have long worked on black holes and, since the 1980s, wormholes — hypothetical tunnels in space that link distant regions of the Universe. About eight years ago, I and my

friend Lynda Obst, a film producer, came up with a movie set on the 'warped side of the Universe' — black holes, wormholes, higher dimensions and beyond. It interested director Steven Spielberg, who brought in Jonathan 'Jonah' Nolan to write the screenplay. Steven dropped out and later Jonah's brother Christopher Nolan took over as director and final screenwriter. Chris and Jonah changed our story almost completely, but preserved the warped space-time and splendidly fulfilled our vision of a science-fiction movie with real science woven deeply in its fabric. In it, nothing violates well-established physical laws and all the wild speculations spring from science, not just the fertile mind of a screenwriter.

How hands-on were you during development?

I met with Jonah and Chris every few weeks as they crafted the screenplay, brainstorming

Interstellar

DIRECTED BY CHRISTOPHER NOLAN
Warner Brothers: 2014.

The Science of Interstellar

KIP THORNE
W. W. Norton: 2014.

about the science. I worked on the visualization of black holes and wormholes with Oscar-winner Paul Franklin and his team at Double Negative Visual Effects in London.

Black holes do not emit light, so you visualize them through gravitational lensing — how they bend light from other objects. I took equations based on Einstein's general theory of relativity and created a description of a wormhole with three parameters: diameter, interior length and the degree of flare where the wormhole joins the external Universe. Paul's team used my equations to compute what a camera would see through the wormhole; Chris, perusing the images, chose the parameter values for *Interstellar's* wormhole.

Did you learn anything new?

With computing power beyond what is normally used by physicists, and software designed to give rapidly changing images at IMAX resolution, we were able to see something never seen before. We simulated a fast-spinning black hole and a field of stars, and what we discovered is an amazingly complex,

ILLUSTRATION BY JIM SPENCER

fingerprint-like pattern of starlight near the edge of the black hole's shadow, which comes from this gravitational lensing. There are regions where the stars look as if they are still, right next to others where the stars are swirling around. When we first saw hints of this, I thought we had done something wrong. We now think it is caused by a complex set of caustics (wrinkles) in the camera's 'past light cone' — not unlike the patterns on the bottom of a sunlit swimming pool. These wrap around the sky many times because the black hole's spin makes space whirl.

Was there any culture clash?

No, there was a full embrace of this melding of arts with science that extended to all four lead actors. Matthew McConaughey and Anne Hathaway came to me for in-depth discussions — they were trying to wrap their heads around the science. Michael Caine asked to have his photograph taken with me, and my jaw dropped. He told me his character was based on me and he wanted to talk about how a theoretical physicist thinks. Jessica Chastain asked for help with quantum equations. The nicest thing was working with artists who are brilliant, intellectually curious and from a background so different from my own.

Will interstellar travel ever be possible?

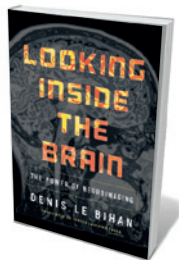
The nearest potentially habitable planet outside the Solar System is perhaps just under 12 light years (3.7 parsecs) away, orbiting the star τ Ceti. If you think of that distance as like going from New York to Perth, Australia, then the distance from Earth to the Moon would be about 7 centimetres. That gives you some sense of the challenge involved. I think humans will make that journey, but not in this century or the next, or maybe the one after that. It is too hard. For a science-fiction story, a wormhole created by an advanced civilization is the only way to do it in the next century, but it is unlikely that wormholes exist. You have to prop them open with 'negative energy' and it is unlikely that the laws of physics allow you to collect enough negative energy. But there is no proof that they can't exist.

What areas of physics currently excite you?

My passion is to understand the non-linear dynamics of warped space-time, and the ideal venue for this is black-hole collisions. There is a high probability that in the next several years we will detect gravitational waves — ripples in the fabric of space-time — generated by such collisions. A combination of computer simulations and gravitational-wave observations will really open our eyes about the behaviour of warped space and warped time when they are wildly dynamical. And who knows, maybe the next movie will involve colliding black holes. We will have to see! ■

INTERVIEW BY ZEEYA MERALI

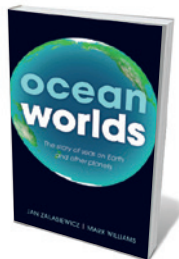
Books in brief



Looking Inside the Brain: The Power of Neuroimaging

Denis Le Bihan (Translated by Teresa Lavender Fagan) PRINCETON UNIVERSITY PRESS (2014)

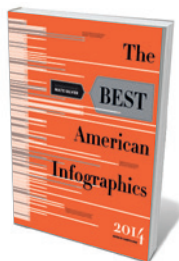
Who better to lead us into the technological wonderland of *in vivo* brain imaging than Denis Le Bihan, pioneer of diffusion magnetic resonance imaging (dMRI)? His densely factual narrative takes us from the seminal work of French surgeon Paul Broca through evolving technologies such as X-rays and MRI. Now the story is being carried forward by Le Bihan's brainchild NeuroSpin, a French Atomic Energy Commission institute for ultra-high-field brain imaging. In 2015 it will gain an 11.7-tesla magnet and begin the search for a 'neural code'.



Ocean Worlds: The Story of Seas on Earth and Other Planets

Jan Zalasiewicz and Mark Williams OXFORD UNIVERSITY PRESS (2014)

For this exhilarating foray into our increasingly water-rich Universe, geologist Jan Zalasiewicz and palaeontologist Mark Williams navigate oceans earthly and elsewhere. They explore our ocean systems' formation in deep time, eons of service as life's nursery, current state of crisis and evaporation a billion years hence, under a Sun that will be 10% hotter. They then boldly go to seas actual and putative in the Solar System and beyond, such as the methane-rich Kraken Mare on Saturn's moon Titan, or oceans on exoplanet 55 Cancri e, which may host expanses of that ambiguity, pressurized 'supercritical' water.



The Best American Infographics 2014

Edited by Gareth Cook MARINER (2014)

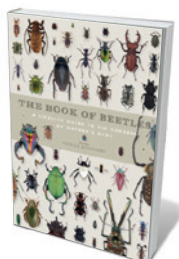
The second in the US infographics series edited by Pulitzer-prizewinning journalist Gareth Cook, this is as brilliant as its predecessor. Try the 'Taste Tube Map': James Wannerton's "synaesthetic tour" ascribing mouthfeel to the name of every London Underground station: Bethnal Green is "Boiled Cabbage"; Mile End, "Fingernails". *National Geographic's* 'In Harm's Way' is a gorgeous rendering of how tornadoic thunderstorms form. And *Nature's* own 'In the Flesh' beautifully realizes research suggesting that *Tyrannosaurus rex* may have sported fetching plumage.



Amnesia

Peter Carey FABER & FABER (2014)

This tale of cyber-hackery set against recent upheavals in Australian government is audacious and deliriously strange. Self-described "shit-stirrer" Felix Moore, a journalist in meltdown, is commissioned to write an exclusive on Melbourne radical Gaby Baillieux. Through Moore's eyes and taped conversations, Gaby's relationship with proto-geek Frederic Matovic emerges as a fury-fuelled mission to right corporate wrongs. The two, with thousands of others, are eventually, paradoxically, liberated by malware. Peter Carey's novel jolts us out of a collective amnesia about political trouble Down Under.



The Book of Beetles: A Life-Size Guide to Six Hundred of Nature's Gems

Patrice Bouchard IVY (2014)

A beetle with "dense, decumbent golden pubescence" inhabiting an air bubble in submerged wood: what's not to love? Imaged in glorious, furry close-up, *Lutrochus germari* is just one of the 600 species (out of a total of 400,000) featured in this paean to the insects that entranced evolutionary-theory pioneer Alfred Russel Wallace. Curator Patrice Bouchard has picked a bevy of beauties, from the tubercle-covered *Gagatophorus draco* to the stunning Atlas beetle (*Chalcosoma atlas*) — three-horned, lustrous and the size of a human hand. *Barbara Kiser*

Correspondence

Let poor countries into rich research

Researchers and decision-makers from low-income countries should help to define the research priorities of universities in high-income countries (see *Nature* **514**, 287; 2014). A strong guild of academics will aid the development of local solutions in poor countries that have little resilience to climate change, biodiversity loss or global health issues (see, for instance, C. Lachat *et al.* *PLoS Med.* **11**, e1001593; 2014).

It is all too common for Western scientists to engage in research in the developing world that fails to benefit their locally based peers (see K. Chu *et al.* *PLoS Med.* **11**, e1001612; 2014). This inequity could be rectified by introducing and enforcing strict ethical standards for funding and research. More-equitable metrics for research evaluation should also be developed (see, for example, E. Smith *et al.* *BMC Med. Ethics* **15**, 42; 2014).

Carl Lachat Ghent University, Belgium.

Patrick Kolsteren Ghent University, Belgium; and Institute for Tropical Medicine, Antwerp, Belgium.

Dominique Roberfroid Institute for Tropical Medicine, Antwerp, Belgium.

carl.lachat@ugent.be

Better modelling for the energy mix

Contrary to Haewon McJeon and colleagues' implications (*Nature* **514**, 482–485; 2014), natural shale gas and renewables are not necessarily competitors. They are driven by different policies and markets.

Over the next decade or so, natural gas can reduce coal consumption and can be rapidly dispatched to balance the intermittency of energy from wind and solar sources. In the longer term, say over the next

50 years, this mix can provide a path to an affordable, low-carbon energy system, especially if it is configured as a backup or coupled with energy storage.

The low capital and high operating costs for natural gas combine well with the high capital and low operating costs of wind and solar. And by reducing the cost of climate mitigation, natural gas should motivate more-aggressive climate policies by governments. Gas is therefore promising because it is a powerful climate-mitigation option, not because it will prevent climate change in the absence of policy.

Moreover, the decisions that shape global energy choices are not always economic or technical (see B. K. Sovacool *Nature* **511**, 529–530; 2014). Effective modelling needs to capture all of the complex and often non-quantitative forces that determine outcomes. These include energy security, trade policy, social equity, weather, distribution of resources, dynamic policies and unforeseen technological advances. Without these factors, the reliability of the long-term forecasts from McJeon *et al.* seems low.

Alexander Q. Gilbert Haynes and Boone, Washington DC, USA.

Benjamin K. Sovacool Aarhus University, Denmark.
benjaminso@hih.au.dk

Funding plea for rural lab outreach

The Imagine Project, set up last year by a group of Brazilian researchers (see go.nature.com/yndqs1), is taking scientific knowledge out of the laboratory to share it with people from rural and indigenous communities. Despite their remarkable response, the initiative is still struggling to be funded.

Scientists and students from universities in Brazil, Morocco, Angola, Mexico and Peru are teaching, for example, molecular biology to Brazilian Indians, genome analysis to Peruvian Inca

descendants and facts about solar energy to teenage land-workers (see go.nature.com/fdqnws).

Community members work and study on alternate days. The learners have, without exception, shown great enthusiasm. Their manual skills often exceed those of urban university students.

The project has been totally funded by Brazil's Federal University of Santa Catarina since its inception. Applications to corporate and governmental funding organizations have so far met with a dispiriting response.

The knowledge imparted is viewed as “too universal” — with no potential for immediate conversion into solutions for food, health or environmental problems. Policy-makers seem to regard the people who are engaged as too remote, too few and too ‘invisible’ to justify involving them in modern developments in science and technology.

André Ramos, Guilherme Razzera Federal University of Santa Catarina, Florianópolis, Brazil.

andre.amos@ufsc.br

Approximation aids handling of big data

We need a radical shift in our approach to data analysis — towards approximation. Our technical capacity is being overtaken by the unprecedented rate of data generation by today's powerful instruments and computers (see H. Esmaeilzadeh *et al.* *Commun. ACM* **56**, 93–102; 2013).

In my view, it is often more efficient to replace conventional, precise computations with scalable approximations that have tight error bounds — in algorithms and data structures, for example. This should be adequate for interpreting data, particularly when refining the explorative phase of an analysis. Researchers may ultimately

still want to perform detailed analyses as indicated trends become clearer. And for some applications, approximation will not work at all.

Approximations, however, are already accepted in many realms, such as next-generation DNA sequencing (M. L. Metzker *Nature Rev. Genet.* **11**, 31–46; 2010). It is time to let mathematical ingenuity replace our obsession with precision.

Thomas Heinis Imperial College London, UK.

t.heinis@imperial.ac.uk

EC grant applicants need fuller feedback

The commitment of the European Commission (EC) to equitable and rigorous scientific review of the research it funds is undermined by the perfunctory quality of feedback to grant applicants. This denies investigators the opportunity to learn and improve from constructive expert criticism.

The commission's review procedures are time-consuming and costly. So, regardless of the final decision on funding, individual applicants should be privy to the full text of expert critiques of their proposals — not merely to a brief summary of the consensus ruling. The anonymous reviewers also deserve to be heard.

In my experience, these evaluation summary reports contain only vague comments, without constructive suggestions or scientific advice. For grant applications that can run to 70 pages, this minimal feedback seems inadequate.

The European Union cannot afford the current waste of intellectual resources and input into scientific development.

Danuta Wasserman European Psychiatric Association, Strasbourg, France; and Karolinska Institute, Stockholm, Sweden.

danuta.wasserman@ki.se

FORUM Depression

The best way forward

Conventional behavioural mouse models of depression are often used to study the disorder, but cannot capture the full picture of the human disease. Here, scientists present two views about the best research strategies to adopt if treatments are to be improved.

THE TOPIC IN BRIEF

- Depression is one of the most prevalent disorders of mental health.
- The development of drugs to treat depression has been stalled for decades.
- Some think that a biological understanding of the mechanisms by which known, rapidly

acting antidepressants exert their effects will facilitate the development of improved treatments.

- Others believe that elucidating the neuronal circuits that mediate the symptoms of depression is paramount if the field is to move forward.

Put therapies first

LISA M. MONTEGGIA

Current antidepressant medications work by increasing the levels of monoamine neurotransmitter molecules at synapses (the junctions between neurons across which chemical signals are transmitted). Although monoamine levels change quickly once treatment with antidepressants starts, the drugs typically take several weeks to exert a clinical effect. Moreover, although they are effective in some patients, in others — notably, those most at risk of suicide — there is often no response. This emphasizes the pressing need to develop rapidly acting antidepressant medications with limited side effects. Many researchers are trying to do this by modelling depression-like characteristics in animals, with the aim of identifying the neuronal circuits underlying the disorder. However, I believe that a better way to design improved drugs is to elucidate the biological mechanisms that underpin effective, rapidly acting antidepressants.

Studies of depression using animals largely focus either on behavioural tests such as stress responses, or on attempts to model and measure aspects of the disorder, such as helplessness or anhedonia (the inability to experience pleasure). Rodents that show intrinsic depression-like behaviour are also studied¹. These models should be treated with caution², because it is unclear how well they correspond to the human behaviours they are purported to replicate. Nonetheless, their use

has been undeniably valuable for developing a constructive debate about the abnormal changes in physiology (the pathophysiology) that lead to depression and, by extension, about treatment options.

But despite the positive aspects of animal models, the most direct route to improving treatments — understanding the biology underpinning therapies that have some success in humans — is surely the best way forward. This approach has proved successful in drug-addiction studies. These have focused on the biological action of cocaine, yielding insights not only into potential therapies but also into the neuronal circuits that underlie addiction. But it is difficult to determine how classical antidepressants work because of the time it takes for the drugs to exert their effects. The discovery that acute, low-dose ketamine triggers a rapid antidepressant response³, even in patients categorized as treatment-resistant⁴, provides a golden opportunity to better delineate the pathways involved in antidepressant action.

Since the clinical observations regarding ketamine were made^{3,4}, their underlying biological mechanisms have been teased apart in animals in the laboratory — an approach known as reverse translation (Fig. 1a). This has led to the identification⁵ of ketamine's target (a receptor protein for the neurotransmitter molecule glutamate) and to the finding that blockade of this receptor by ketamine is linked to changes in a specific signalling pathway. These changes may result in the increase in synaptic efficiency that

underlies the drug's rapid effects⁶.

Such findings move us closer to understanding how ketamine acts, and may help to identify drugs that work in similar ways. In particular, the knowledge gleaned from this reverse-translation approach indicates that an antidepressant response can be triggered in a manner distinct from those of classical antidepressant drugs. The discovery finally moves the field beyond the old focus on increasing synaptic monoamine levels⁶.

There are clear problems with using animals to model neuropsychiatric disorders, because such conditions are probably caused by a variety of factors. But if the biological basis is to be understood, animal models will obviously be required to examine the genetic variants and genetically linked characteristics associated with complex disorders such as depression. However, no animal model will ever fully correspond to this devastating human condition. Therefore, such efforts must be combined with the investigation of promising pharmacological leads in humans if improved treatments are at last to be found.

Lisa M. Monteggia is in the Department of Neuroscience, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9111, USA.
e-mail: lisa.monteggia@utsouthwestern.edu

Fix faulty circuits

ROBERT C. MALENKA & KARL DEISSEROTH

Drugs to treat psychiatric disorders such as depression were discovered serendipitously in the 1950s, concomitant with the realization that drugs can also cause syndromes resembling mental illness⁷. The ensuing decades saw attempts to identify the molecular targets of antidepressants and the evolution of simple biochemical models of mental illness. Such pharmacological approaches have provided minimal insight into the

DEPRESSION

A Nature special issue
nature.com/depression

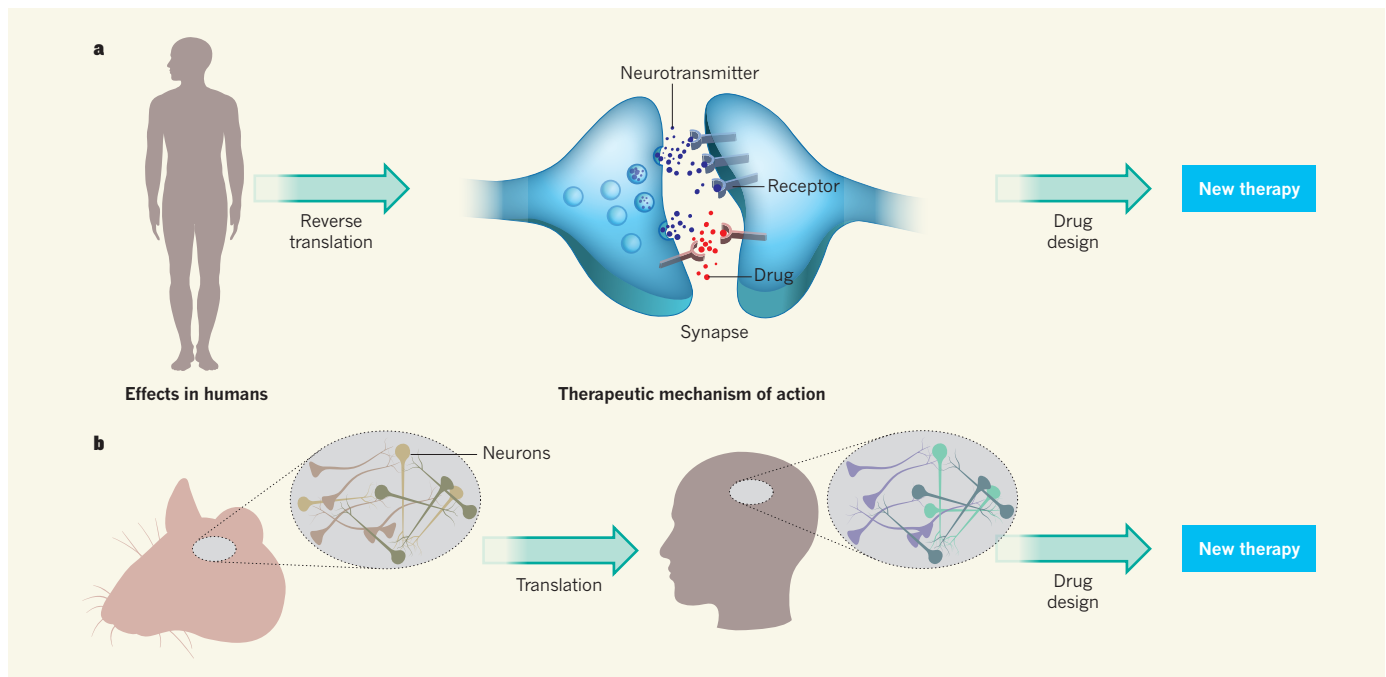


Figure 1 | Approaches to drug discovery for depression. There is a consensus among researchers that successful drug development will require a concerted effort, embracing approaches based on both pharmacology and neuronal circuits. However, there is debate about where such an effort should focus first. **a**, Antidepressants that are effective in humans can be reverse translated in the laboratory, to infer how treatments can be improved. The mechanisms of action of effective drugs are teased apart, to give a biochemical understanding of how an antidepressant such as ketamine exerts its effects at

synapses, the neuronal junctions across which neurotransmitter molecules are passed. This knowledge can then be used to identify new targets for treatment or to develop more-effective drugs that act on known targets. **b**, An alternative approach is to dissect the neuronal circuitry that malfunctions in animals showing symptoms of depression. Once a dysfunctional circuit has been identified, the same circuit can be analysed in humans. In this way, putative drug targets can be defined and tested in a reliable animal model, and then eventually tested in humans.

pathophysiology of mental illness, and have led only to the development of drugs that have similar biochemical actions to those of their predecessors, rather than improved efficacy⁷. However, this strategy is still commonly taken by both academics and pharmaceutical companies. We believe that a sophisticated understanding of the pathophysiology of mental illness at the level of neuronal circuits provides a more rational way forward.

A prototypical example of the pharmacological approach is a model of schizophrenia based on administering ketamine to animals and humans⁸. It could be argued that studying how ketamine acts will not only help us to understand schizophrenia better, but might also provide clues to treating depression, because it is an effective short-term treatment for this disorder. Although we welcome such pioneering clinical work, the decades-old history of psychiatry research indicates that attempts to define this drug's therapeutic mechanism of action are doomed to failure⁷. Those who cannot remember the past will almost certainly repeat it.

In its broad action, ketamine treatment is analogous to electroconvulsive therapy (ECT) and lithium. These are effective therapies for refractory depression and bipolar disorder, respectively, but, like ketamine, affect neurons indiscriminately throughout the brain. As a result, the many attempts to define the biological mechanisms by which ECT

and lithium act have led neither to a better understanding of the changes in physiology that are responsible for disease nor to improved treatments.

The path to better treatments might more suitably lie in defining the specific brain circuits that mediate symptoms of mental illness. Improved methods for recording and manipulating neural activity have led to advances in defining the circuits that mediate several symptoms. For instance, in rodents, manipulation of key cells in the brain's reward circuitry has profound effects on anhedonia^{9,10}. Similarly, circuits involving another area of the brain have been found to mediate features of anxiety, which often accompanies depression^{9,10}.

Although conventional animal models of psychiatric symptoms (which are based on pharmacology) have legitimately been called into question², advances in our understanding of the genetics of mental illness have facilitated the development of rodent models with strong construct validity — that is, behavioural abnormalities in the model are brought about by the genetic defects that cause disease in humans. These models provide ideal systems in which to dissect the circuitry that mediates maladaptive behaviours. Importantly, circuits found to be malfunctioning in animals can be tested in human brains using imaging and stimulation techniques (Fig. 1b). In this way, researchers are beginning to identify the circuit dysfunctions that

cause key symptoms of mental illness^{9–11}.

This process may ultimately allow drug development to become rational. The aim is to identify targets in neuronal circuits that, when manipulated, can repair the dysfunction. After more than 50 years of stagnation, the field may be ready to move beyond the same pharmacological approaches and instead leverage more-suitable methods for advancing our understanding of the circuitry involved¹¹. It is time to make history, not repeat it. ■

Robert C. Malenka and Karl Deisseroth are in the Departments of Psychiatry and Behavioral Sciences and of Bioengineering, Stanford University, Stanford, California 94305, USA.
e-mail: malenka@stanford.edu

- Li, B. *et al. Nature* **470**, 535–539 (2011).
- Nestler, E. J. & Hyman, S. E. *Nature Neurosci.* **13**, 1161–1169 (2010).
- Berman, R. M. *et al. Biol. Psychiatry* **47**, 351–354 (2000).
- Price, R. B., Nock, M. K., Charney, D. S. & Mathew, S. J. *Biol. Psychiatry* **66**, 522–526 (2009).
- Autry, A. E. *et al. Nature* **475**, 91–95 (2011).
- Kavalali, E. T. & Monteggia, L. M. *Am. J. Psychiatry* **169**, 1150–1156 (2012).
- Hyman, S. E. *Neuropsychopharmacology* **39**, 220–229 (2014).
- Moghaddam, B. & Krystal, J. H. *Schizophrenia Bull.* **38**, 942–949 (2012).
- Deisseroth, K. *Nature* **505**, 309–317 (2014).
- Steinberg, E. E., Christoffel, D. J., Deisseroth, K. & Malenka, R. C. *Curr. Opin. Neurobiol.* **30**, 9–16 (2015).
- Akil, H. *et al. Science* **327**, 1580–1581 (2010).

Magnetic fields without magnetic fields

Exquisite control of quantum systems has allowed researchers to connect reality to ideas of how an exotic form of particle transport known as the quantum Hall effect can occur in the absence of a magnetic field. [SEE LETTERS P.237 & P.241](#)

JONATHAN SIMON

A magnetic field produces a force on a charged particle that is perpendicular to both the field and the particle's velocity. This force, called the Lorentz force, induces circular orbits, the chirality (handedness) of which breaks time-reversal symmetry, giving rise to unusual particle-transport phenomena such as 'edge states' protected by topological properties and the related quantum Hall effect (Fig. 1a). In two papers published in this issue of *Nature*, researchers have experimentally demonstrated that the quantum Hall effect can persist in the absence of a magnetic field (Jotzu *et al.*¹, page 237), and have developed a new method to quantify topological protection in arbitrary dynamical systems (Roushan *et al.*², page 241).

In 1988, the physicist F. Duncan Haldane introduced a model that exhibited the quantum Hall effect without a net magnetic field, instead using a field that alternated within each unit cell of a honeycomb lattice³. Although his model has never been observed

in the laboratory, it forced researchers to ask what ingredients are absolutely necessary for the quantum Hall effect and how they could predict whether its signatures will be present in a particular material. At its heart, the quantum Hall effect is a voltage induced across a piece of material (the 'Hall bar') when a current flows along the material; the voltage is proportional to small magnetic fields, and quantized to discrete values for large magnetic fields. That such a quantized voltage could persist in the absence of a magnetic field was quite unexpected. We now understand that the key ingredient is not a magnetic field, but rather Berry curvature⁴, a mathematical entity that smoothly connects quantum states of different momenta and acts like a generalized magnetic field. These deep ideas are at last connected to experimental reality: Jotzu *et al.* implemented and studied Haldane's model using cold atoms (Fig. 1b), whereas Roushan *et al.* used a superconducting circuit to measure the Berry curvature in systems of one and two quantum bits (qubits; Fig. 1c).

The first step to realizing Haldane's model is

creating a honeycomb lattice, which Jotzu and colleagues previously demonstrated for potassium atoms trapped by laser beams⁵. Their new breakthrough is a synthetic magnetic field for the atoms, which, being charge neutral, do not experience a Lorentz force from an actual magnetic field. The key is to generate a large synthetic magnetic flux, which is the number of field lines piercing an area. Although small amounts of flux had previously been generated⁶, the large flux necessary for Haldane's model had not. Jotzu *et al.* show that circular shaking of individual sites of a honeycomb lattice results in potassium atoms hopping to non-adjacent lattice sites (Fig. 1b) in a way that is equivalent to a large, alternating flux. The idea of generating chiral motion of quantum particles by shaking is a modern development^{7,8}, and, before Jotzu and colleagues' work, had been definitively observed only in bismuth selenide (Bi₂Se₃) samples subjected to short-duration laser pulses⁹.

Jotzu *et al.* measured the presence of their synthetic magnetic flux by accelerating the potassium atoms and searching for a force perpendicular to their motion, akin to the Lorentz force for electrons in magnetic fields. They observed this perpendicular force as motion of the atoms perpendicular to their velocities. Crucially, they found that the direction of the force changed depending on the direction of atomic velocity, indicating a time-reversal-breaking force, because a time-reversal-symmetric force would be in the same direction regardless of the particle's velocity. Finally, the authors mapped out the sensitivity of Haldane's model to disorder in the lattice-site energies, demonstrating that such perpendicular motion persists only when the

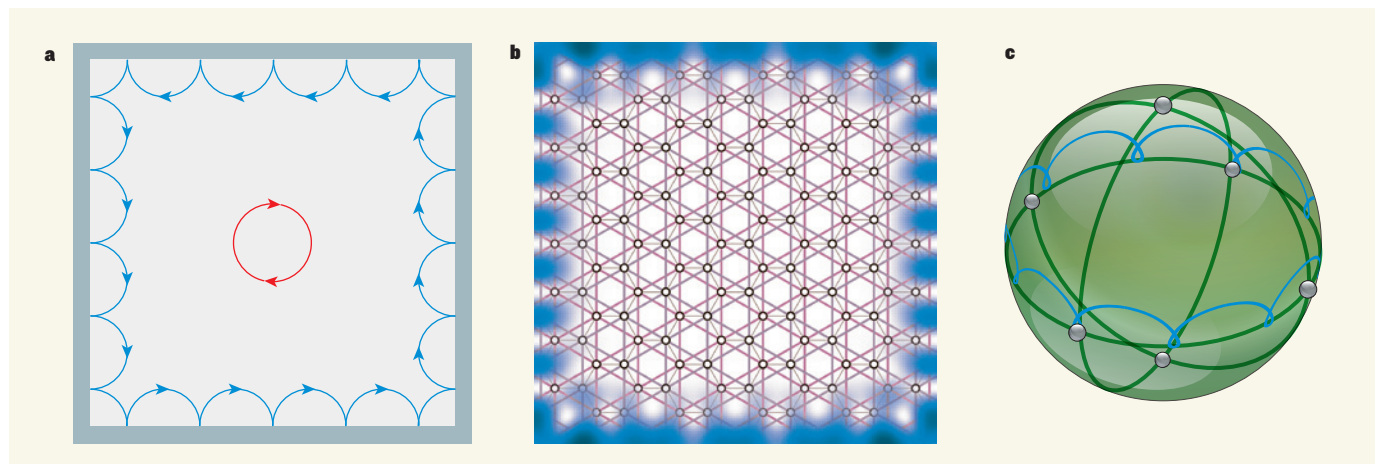


Figure 1 | Chiral dynamics. **a**, A charged particle in a magnetic field oriented perpendicularly to the page undergoes circular orbits (red) of specific chirality in the bulk of a two-dimensional material. If it encounters an edge, it moves in a skipping orbit along that edge (blue). **b**, In Haldane's model, which Jotzu *et al.*¹ implemented, a particle moves in a honeycomb lattice (black sites) with non-adjacent neighbour particle tunnelling (magenta and purple; adjacent tunnelling shown in grey) arranged to produce a magnetic field that alternates within each unit cell. In spite of the absence of any net field within a unit cell, a chiral edge channel of particle motion (blue) such as the one in **a** persists.

c, Roushan *et al.*² studied the chiral dynamics of a quantum bit (qubit), the state of which (grey dots) can be visualized on the surface of a sphere, where the top of the sphere is state '0' and the bottom is state '1'; quantum superposition states of '0' and '1' states reside in between, with the azimuthal angle determined by the phase of the superposition state, and the polar angle determined by the relative contributions of states 1 and 0 (green lines). The authors prepared the qubit in a particular state and manipulated it in such a way that the state began to wobble (blue). The form and magnitude of the wobbling allowed them to extract the Berry curvature, a quantity that characterizes the system's chiral dynamics.

shaking produces enough flux to overcome the disorder imposed for testing purposes.

Meanwhile, Roushan *et al.* demonstrated a technique for measuring Berry curvature in arbitrary quantum systems, generalizing the perpendicular-motion signature used by Jotzu and co-workers: for slow variations of physical parameters such as momenta or magnetic fields, a quantum system can remain in its ground state up to a Lorentz-force-like correction to its wavefunction that is proportional to both the rate of change of parameters and the Berry curvature. Imagine transporting a brimming bowl of soup: move it slowly and it will not spill, but sudden movements make a mess. Now, if stumbling forward made the soup spill to the right, the apparent force that the soup experienced could be understood in terms of Berry curvature. By tuning the parameters of their quantum systems and measuring deviations of the quantum state from the ground state, Roushan *et al.* extracted the Berry curvature at each point in the mathematical space formed by the parameters of the system¹⁰.

They first implemented the simplest possible quantum system — a single qubit with an adjustable energy difference between its two states, along with a tunable microwave field driving the qubit from one state to the other. Formally, this system is equivalent to a particle of known momentum in a honeycomb lattice, where the qubit state reflects the probability of the particle being on each lattice site within a unit cell. By varying the energy difference and the microwave field, Roushan *et al.* dragged the qubit through momentum space and watched for deviations, not in position as in Jotzu and colleagues' study, but in qubit state. Exquisite control of the system let them precisely measure the quantum state at any instant, and determine the Berry curvature.

But Roushan and colleagues went a step further: they studied a system of two interacting qubits and mapped out its Berry curvature. This is a stimulating and wonderful idea because the two-qubit system is not analogous to particles on a lattice, but the Berry curvature remains well defined and non-zero. We must then reconsider what the Berry curvature means — it is not just about particles moving in circular orbits, but more subtly quantifies global topological properties of arbitrary quantum systems that are insensitive to local disorder. Roushan and co-workers' technique is potentially scalable to systems containing many qubits, suggesting that quantum magnets and even quantum computers could be understood in terms of Berry curvature and topological properties. Nonetheless, the number of measurements required to scale up is daunting, and the connection between Berry curvature and dynamics in the high-dimensional parameter spaces associated with such systems remains an active field of research.

Meanwhile, Jotzu and colleagues' work is a crucial step towards cold-atom realizations of

exotic phenomena such as fractional quantum Hall phases and fractional Chern insulators. These are both ordered arrangements of electrons in which imperfections in the ordering, known as anyons, behave as particles with a bizarre property — they remember their past locations and can be braided around one another, becoming entangled like shoelaces. These anyons would be fascinating to observe in their own right, and could be a powerful paradigm for quantum computing.

All that remains for Jotzu and co-workers is to add interactions between their potassium atoms and to harness the extraordinary control of the cold-atom toolbox to search for these phenomena. Nonetheless, these phenomena exist near the ground state of ultracold atoms interacting with one another in the presence of a magnetic field, and reaching that ground state requires cooling to one-billionth of a degree above absolute zero, something we do not know how to do for general quantum systems. The

next challenge is thus to find a way to smoothly convert the lowest-temperature state of matter yet created, the Bose–Einstein condensate, into a fractional quantum Hall phase, or to make a state-of-the-art refrigerator¹¹. ■

Jonathan Simon is in the Department of Physics, University of Chicago, Chicago, Illinois 60637, USA.

e-mail: simonjon@uchicago.edu

1. Jotzu, G. *et al.* *Nature* **515**, 237–240 (2014).
2. Roushan, P. *et al.* *Nature* **515**, 241–244 (2014).
3. Haldane, F. D. M. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).
4. Berry, M. V. *Proc. R. Soc. Lond.* **392**, 45–57 (1984).
5. Tarruell, L. *et al.* *Nature* **483**, 302–305 (2012).
6. Lin, Y. J. *et al.* *Nature* **462**, 628–632 (2009).
7. Lindner, N. H., Refael, G. & Galitski, V. *Nature Phys.* **7**, 490–495 (2011).
8. Kitagawa, T. *et al.* *Phys. Rev. B* **82**, 235114 (2010).
9. Wang, Y. H. *et al.* *Science* **342**, 453–457 (2013).
10. Gritsev, V. & Polkovnikov, A. *Proc. Natl Acad. Sci. USA* **109**, 6457–6462 (2012).
11. Kapit, E., Chalker, J. T. & Simon, S. H. Preprint at <http://arxiv.org/abs/1408.0959> (2014).

PALAEONTOLOGY

Mystery of the horrible hands solved

A pair of newly discovered 70-million-year-old fossils from Mongolia — including material previously lost to poaching — reveals the true nature of one of the most enigmatic dinosaur species, *Deinocheirus mirificus*. SEE LETTER P.257

THOMAS R. HOLTZ JR

Although some species of fossil vertebrate are known from hundreds or even thousands of individual skeletons, most are known from the incomplete remains of a single individual, making it difficult to infer the organism's anatomy, ecology and evolutionary origins. This is especially troubling when the fragmentary remains point to animals of unusual appearance. Thankfully, continued fieldwork can lead to the discovery of additional remains and help to fill the gaps in our knowledge. Such is the case for *Deinocheirus mirificus*, a dinosaur found in the Nemegt Formation in Mongolia, which dates to around 70 million years ago, during the Late Cretaceous period. *Deinocheirus* was known only from a gigantic pair of arms and a few isolated bones — until now. On page 257 of this issue, Lee *et al.*¹ describe two new skeletons of *Deinocheirus*, which add together to give us a nearly complete anatomy of this previously mysterious animal.

In July 1965, Zofia Kielan-Jaworowska of the Polish–Mongolian Palaeontological Expedition discovered the first *Deinocheirus* remains. All that was found were nearly complete

shoulder girdles and forelimbs (the latter an astounding 2.4 metres long, the record-holder for the longest-known arms of a bipedal animal), and a few ribs and vertebrae. These clearly showed the diagnostic traits of a theropod — the group of mostly carnivorous dinosaurs that includes *Allosaurus*, *Tyrannosaurus* and *Velociraptor*, and also, through descent, modern birds. Despite their incompleteness, the remains were sufficiently distinct in size and morphology from other known theropods to allow Kielan-Jaworowska's colleagues Halska Osmólska and Ewa Roniewicz to propose² that they belonged to a new genus and species, *Deinocheirus mirificus*, or 'unusual horrible hand'.

The lack of a complete skeleton meant that palaeontologists were uncertain as to the place of *Deinocheirus* in the theropod family tree. Although Osmólska and Roniewicz observed some similarities with the smaller and superficially ostrich-like ornithomimosaurs (a group of omnivorous beaked theropods), they remained uncommitted as to a particular affinity with any of the then-known groups of dinosaurs. Some early reconstructions for popular audiences portrayed the dinosaur as a generalized *Allosaurus*-like carnivore with



Figure 1 | A dinosaur emerges. Two new, almost complete skeletons of *Deinocheirus mirificus*, presented by Lee *et al.*¹, have revealed the dinosaur's unusual combination of features.

hypertrophied prey-capturing arms; more bizarrely, others speculated that it was a gigantic sloth-like climber that suspended itself underneath the limbs of (unrecorded in the fossil record) enormous trees³. By 1972, however, palaeontologist John Ostrom had agreed that Osmólska and Roniewicz's initial observations were correct and that, bone for bone, *Deinocheirus* shared several anatomical features uniquely with ornithomimosaurs, and so was most likely an 'ostrich dinosaur' of hitherto unrecognized gigantic proportions⁴. Although some have been sceptical of this claim⁵, phylogenetic analyses have supported the hypothesis that *Deinocheirus* was a primitive ornithomimosaur⁶, and most modern restorations of this dinosaur have portrayed it as such.

Lee and colleagues' discoveries show that *Deinocheirus* in fact had a series of unusual features not seen in combination before (Fig. 1). The dinosaur's toothless snout flares out to the sides, similarly to that of its contemporaries the hadrosaurids (duck-billed herbivorous dinosaurs). Its lower jaw is particularly massive for an ornithomimosaur, more reminiscent of other large-bodied theropods. The neural spines of the lower back, hips and base of the tail are extended vertically to form a tall sail,

somewhat like that of the distantly related *Spinosaurus*. And *Deinocheirus* retains the relatively short and broad foot of a typical theropod, unlike the long, slender feet of the ornithomimid ornithomimosaurs (such as its smaller contemporary *Gallimimus*) and of the tyrannosaurids (such as the predator of *Deinocheirus*, *Tarbosaurus*⁷). The toes end in unusually squared-off hooves.

Because nearly the entire skeleton is now known, the evolutionary position of *Deinocheirus* can be estimated much more securely. It is indeed an ornithomimosaur, and is placed by Lee *et al.* with the earlier *Garudimimus* and *Beishanlong* (previously the largest known ornithomimosaur) to form a group to which Osmólska and Roniewicz's name Deinocheiridae has been applied. Both deinocheirids and the smaller ornithomimids lack teeth (more-primitive ornithomimosaurs have simple peg-like teeth), but whereas the ornithomimids evolved numerous adaptations for increased speed, the deinocheirids seem to have specialized for size and weight bearing.

Where did *Deinocheirus* fit in the ecology of the Nemegt Formation — an environment similar⁸ to that of today's Okavango Delta, which drains into the Kalahari desert? The

anatomy of ornithomimosaurs (including *Deinocheirus*) is consistent with a diet of plants and small animals⁹; in addition, Lee *et al.* document the presence of fish remains in the belly of one of their specimens. Thus, *Deinocheirus* was seemingly a 6.4-tonne omnivore, inhabiting an ecosystem with several gigantic potential competitors: the comparably large long-necked herbivorous theropod *Therizinosaurus*; one or more even-larger titanosaurian sauropods (long-necked herbivorous quadrupedal dinosaurs); and the smaller duck-billed *Saurolophus*. These may have competed for trees as a food resource, and the broader diet of *Deinocheirus* may have allowed it to access resources that the other tall herbivores of the community could not. The identification of *Deinocheirus* as a gigantic ornithomimosaur confirms the observation that three separate lineages of non-flesh-eating theropods (ornithomimosaurs, therizinosaurids and the short-beaked bird-like oviraptorosaurs) independently achieved their maximum known size in the later Cretaceous of eastern and central Asia¹⁰.

Alongside the anatomical, evolutionary and ecological implications of this discovery, these particular fossils have an unusual collection history. One of the two skeletons was discovered in 2006 and the other in 2009, as part of the Korea–Mongolia International Dinosaur Expedition, but it was realized in the field that parts of both skeletons had already been poached by illegal collectors. Pascal Godefroit at the Royal Belgian Institute of Natural Sciences informed Lee and colleagues in 2011 that he had seen the skull, hand and feet of *Deinocheirus* in private hands in Europe (see Supplementary Information of the paper¹). In May 2014, these specimens were repatriated to Mongolia and the exact match of the joint surfaces of various bones indicated that the poached remains indeed belonged to these skeletons.

A parallel situation played out recently with another 'mystery' theropod dinosaur. Much of the anatomy of the enormous mid-Cretaceous African fish-eating *Spinosaurus*, known initially from an incomplete skeleton found in 1912¹¹, remained obscure until the publication of a new skeleton and assorted isolated elements of this dinosaur earlier this year¹². As with the *Deinocheirus* material, parts of the skeleton had been poached in the field and were later reunited. The anatomy revealed by these more-complete *Spinosaurus* remains also showed that it was not simply a typical member of its group grown huge, but rather had unsuspected adaptations and perhaps a different mode of life. Both cases reveal that, although inferring the anatomy and ecology of fossil forms from close relatives is often the only reasonable approach, such interpretations can be proved incorrect when primary data are retrieved from the fossil record. ■

Thomas R. Holtz Jr is in the Department of Geology, University of Maryland, College

Park, Maryland 20742, USA, and in the Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington DC, USA.
e-mail: tholtz@umd.edu

1. Lee, Y.-N. *et al.* *Nature* **515**, 257–260 (2014).
2. Osmólska, H. & Roniewicz, E. *Palaeontol. Pol.* **21**, 5–19 (1969).
3. Rozhdestvensky, A. K. *Paleontol. J.* **4**, 117–125 (1970).

4. Ostrom, J. H. *McGraw-Hill Yearbook of Science and Technology for 1971* 176–179 (McGraw-Hill, 1972).
5. Makovicky, P. J., Kobayashi, Y. & Currie, P. J. in *The Dinosauria* 2nd edn (eds Weishampel, D. B., Dodson, P. & Osmólska, H.) 137–150 (Univ. California Press, 2003).
6. Senter, P. J. *Syst. Palaeontol.* **5**, 429–463 (2007).
7. Bell, P. R., Currie, P. J. & Lee, Y.-N. *Cret. Res.* **37**, 186–190 (2012).
8. Jerzykiewicz, T. in *The Age of Dinosaurs in Russia and Mongolia* (eds Benton, M. J., Shishkin, M. A.,

- Unwin, D. M. & Kurochkin, E. N.) 279–296 (Cambridge Univ. Press, 2000).
9. Zanno, L. E. & Makovicky, P. J. *Proc. Natl Acad. Sci. USA* **108**, 232–237 (2010).
10. Zanno, L. E. & Makovicky, P. J. *Proc. R. Soc. B* **280**, 20122526 (2013).
11. Stromer, E. *Abh. König. Bayer. Akad. Wissen. Math.-Phys. Kl.* **28**, 1–32 (1915).
12. Ibrahim, N. *et al.* *Science* **345**, 1613–1616 (2014).

This article was published online on 22 October 2014.

HIGH-TEMPERATURE SUPERCONDUCTIVITY

Electron mirages in an iron salt

The detection of unusual ‘mirage’ energy bands in photoemission spectra of single-atom layers of iron selenide reveals the probable cause of high-temperature superconductivity in these artificial structures. [SEE LETTER P.245](#)

JAN ZAAENEN

Engineering artificial structures on the nanometre scale with the aim of controlling the quantum properties of their electrons is a mainstay of modern materials science. But it is a grand challenge to try to apply this technology to manipulate the rich, collective quantum phenomena of strongly interacting electron systems, because these systems invariably involve complicated chemistry that is hard to tame on the nanoscale. An iconic example of such phenomena is superconductivity (electrical conduction without resistance) in copper- and iron-based metallic salts at unusually high temperatures. Discovered¹ in the 1980s, this type of superconductivity is still poorly understood^{2,3}, and only in the past couple of years has it become possible to manufacture high-quality single-atom layers of an iron selenide superconductor⁴. The surprise is that these nanolayers superconduct at higher temperatures^{4,5} — perhaps⁶ up to a whopping 109 kelvin — than the rather poorly superconducting bulk form of this metallic salt. On page 245 of this issue, Lee *et al.*⁷ report data that seem to uncover the culprit behind this intriguing observation.

The researchers used photoemission spectroscopy to directly probe the electron worlds of iron selenide (FeSe) films, finding in their spectra highly unusual ‘mirage’ energy bands. It turns out that these bands can be explained in terms of physics that is unique to these nanostructures: the special nature of the interaction between the FeSe electrons and phonons (quantized crystal-lattice vibrations) in the electrically insulating strontium titanate (SrTiO₃) substrate on which the films are grown. The interaction seems to provide an ideal way of boosting the superconductivity

of the films. This result is ironic, given the contentious role of electron–phonon interactions in the long history of high-temperature superconductivity⁸.

Conventional superconductivity, which occurs at temperatures close to absolute zero in simple metals such as aluminium, was explained by a theory developed by John Bardeen, Leon Cooper and John Schrieffer in the mid-twentieth century. This theory describes the superconducting phase as a collective quantum state — a Bose–Einstein condensate — of ‘Cooper pairs’ of electrons. Cooper pairs form as a result of a quantum-mechanical exchange of phonons that produces

an attractive interaction between the electrons; on their own, electrons just repel each other.

However, when the news of the discovery¹ of high-temperature superconductivity broke in 1986, it was clear

that phonons could not be responsible for the phenomenon⁸; these are already ineffective at rather low temperatures, and the repulsive interactions between the electrons are exceedingly strong in the metallic salts in which the phenomenon was observed. After nearly 30 years of struggle with intricate quantum many-body theory, eventually a consensus was reached that high-temperature superconductivity can nevertheless arise in these materials. This came about because quantum processes had been identified that could transform the strong electrostatic repulsion between the electrons into effective interactions stabilizing Cooper pairs at high temperature².

The results suggest new routes by which high-temperature superconductors can be engineered on the nanoscale.

This electronic pairing mechanism is very different from that involving phonons^{2,8}, to the extent that the two mechanisms have generically opposing effects in bulk metals. But according to Lee and colleagues’ experiments on FeSe films, phonons in the underlying substrate can work closely together with this electronic pairing to promote superconductivity in the films.

Photoemission spectroscopy measures the probability that an electron of given energy and momentum will be removed from a sample subjected to photon excitation. In simple molecules such as hydrogen (H₂), the nature of the electron–phonon interactions is easily discerned from such spectroscopic information⁹. After an electron has been removed, the molecule becomes an ion (H₂⁺) and so acquires an altered bond length. This is indicated in the photoemission spectrum as a progression of peaks: the lowest-energy peak corresponds to the pure electronic excitation, and is followed by peaks at higher energies that are associated with the emission of one, two, and so on, phonons ‘repairing’ the mismatch in bond length.

Lee *et al.* observed this spectral pattern for the specific case of the electrons in FeSe monolayers on SrTiO₃ substrates. Instead of the lowest-energy peak, the authors detected an energy band that reflects the way in which the electrons quantum-mechanically delocalize in the FeSe, forming the material’s metallic and superconducting states. But the authors also observed intense mirages — replicas of this ‘normal’ band showing up at higher excitation energies. Remarkably, this extra energy is coincident with the energy required to excite dominant phonons of the SrTiO₃ substrate.

Such mirage bands have not been observed in any other solid. The reason is that in bulk metals the electron–phonon interaction becomes short-range owing to ‘metallic screening’ effects. As a result, the electrons excite phonons of all wavelengths, and these phonon ‘shake-offs’ result in featureless backgrounds in the photoemission spectrum. But in the present case, the nanometre-scale structure of the films results in the electrons exciting only phonons of long wavelength in the insulating substrate, because of unscreened interactions in this medium. This wavelength selectivity causes the appearance of mirage bands, and the strength of the mirages indicates that the electron–phonon interaction is strong. What is more, such long-wavelength

phonons actually cooperate with the electronic-pairing mechanism⁷, explaining why the films superconduct at higher temperatures than their bulk analogues.

These are exciting results for superconductivity researchers, because they suggest new routes by which high-temperature superconductors can be engineered on the nanoscale. But perhaps more importantly, they are an example of the kind of surprise that might emerge from studies that attempt to harness the power of nanotechnology to explore the

rich but largely uncharted territory of strongly interacting quantum systems. ■

Jan Zaanen is at the Instituut-Lorentz for Theoretical Physics, Leiden University, 2300 RA Leiden, the Netherlands. e-mail: jan@lorentz.leidenuniv.nl

1. Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189–193 (1986).
2. Keimer, B., Kivelson, S. A., Norman, M. R., Uchida, S. & Zaanen, J. Preprint at <http://arxiv.org/abs/1409.4673> (2014).

3. Steward, G. R. *Rev. Mod. Phys.* **83**, 1589–1652 (2011).
4. Wang, Q.-Y. *et al. Chin. Phys. Lett.* **29**, 037402 (2012).
5. Liu, D. *et al. Nature Commun.* **3**, 931; <http://dx.doi.org/10.1038/ncomms1946> (2012).
6. Ge, J.-F. *et al.* Preprint at <http://arxiv.org/abs/1406.3435> (2014).
7. Lee, J. J. *et al. Nature* **515**, 245–248 (2014).
8. Zaanen, J. in *100 Years of Superconductivity* (eds Rogalla, H. & Kes, P. H.) Ch. 2.4, 92–117 (Chapman & Hall, 2011); <http://arxiv.org/abs/arXiv:1012.5461>.
9. Turner, D. W. *Molecular Photoelectron Spectroscopy* (Wiley, 1970).

NEUROBIOLOGY

Building a bigger brain

An innovative approach to analysing the functions and gene-expression profiles of neural stem cells in developing human and mouse brains sheds light on the differences — and similarities — between the two species. SEE LETTER P.264

FORREST O. GULDEN & NENAD ŠESTAN

The mammalian cerebral neocortex is a thin mantle of neurons on the surface of the brain's hemispheres that is largely responsible for higher-order mental functions. In different mammalian species, the cortex can vary by more than 10,000-fold in mass¹, and such variations have been associated with species-specific differences in cognition and behaviour^{1–4}. Alterations in the proliferation and differentiation of radial glia — non-neuronal progenitor cells^{5–7} that give rise to most of the approximately 16 billion neurons in the human cerebral cortex¹ — are thought to play a crucial part in determining the size of the neocortex. Unfortunately, identifying the genes that regulate development of the radial glia in humans has proved exceedingly difficult. On page 264 of this issue, Lui *et al.*⁸ use a sophisticated approach to identify and characterize the role of a secreted protein, PDGFD, in mediating the proliferation of radial glia in humans.

Radial glia serve both as progenitors and as migratory guides for the daughter cells they give rise to, and so provide for the production and placement of appropriate cell numbers and types in the developing neocortex^{5–7,9}. Changes to division of radial glia, the length of their cell cycle or the rate at which they proliferate may therefore have profound effects on the size, composition and functional repertoire of the neocortex^{2–7}. These cells can be efficiently studied experimentally in mice and many other model organisms, but ethical and practical barriers hinder similar tests in humans. As such, genetic studies have identified many genes that both control proliferation in mice and regulate the size of the human neocortex^{2,10,11}, but the identities of genes involved in

the development of radial glia only in humans or in non-standard laboratory animals have remained elusive.

A complete list of the genes expressed within a sample or cell type, together with information about the abundance of transcripts

produced from each of these genes, is called a transcriptome. When a transcriptome is generated from a complex biological tissue such as the neocortex, it is an amalgamation of the transcriptomes of the millions or billions of cells that make up that tissue. However, cells of the same type and developmental stage tend to have more-similar transcriptomes than those of different cell types and, furthermore, cells of a given type are generally not distributed evenly throughout a tissue. Thus, it stands to reason that the expression levels of all of the genes expressed by a given cell type will rise and fall in unison as the abundance of that cell type varies in different regions of the tissue under analysis. This provides the basis for identifying the transcriptome of a single

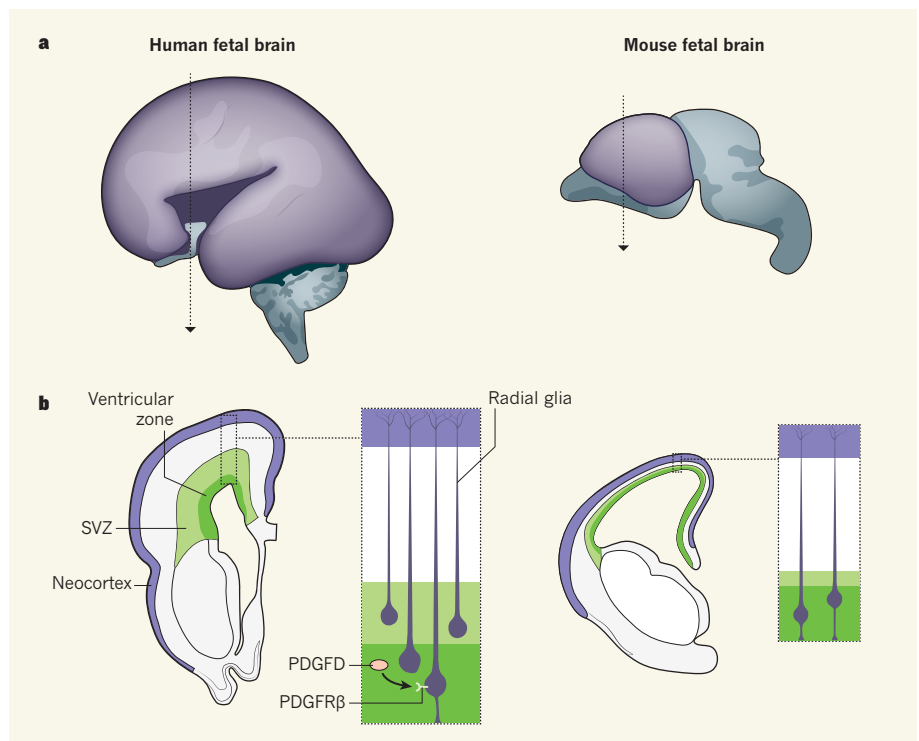


Figure 1 | Transcriptional control in radial glia. **a**, Side-on views of human and mouse brains at equivalent developmental stages (not to scale). The developing cerebral neocortex is indicated in purple. **b**, A cross-section of one brain hemisphere from each species, taken from the dotted lines shown in **a**. The magnified sections show the distribution of cells in the brain called radial glia. Lui *et al.*⁸ report that the protein PDGFD, acting through the receptor protein PDGFRβ, signals to the radial glia of humans but not to those of mice. PDGFD signalling correlates with increases both in proliferation of radial glia and in the distribution of cells resembling radial glia throughout the ventricular zone and subventricular zone (SVZ).

cell type, even when starting with a mixed population of cells^{12,13}.

Following this reasoning, the authors generated transcriptomes from 87 cross-sections of tissue taken from a single human fetal neocortex. From these data, they identified 55 groups of genes, called modules, in which gene expression rose and fell together, reflecting differences in the cellular composition of each cross-section. Six of these modules contained genes that are already known to be expressed by radial glia in mice, suggesting that these modules each represented a portion of the transcriptome of human radial glia. Importantly, these six modules contained several genes that the researchers found to be expressed in human, but not murine, radial glia. Lui and colleagues chose to focus on just one of these genes, *PDGFD*.

This gene encodes a secreted growth factor that has not previously been implicated in neocortical development¹⁴. The authors report that inhibiting PDGFD protein signalling in slices taken from the human neocortex reduced the proliferation of radial glia. Conversely, introducing abnormally high levels of either PDGFD or a permanently activated form of its receptor protein, PDGFR β , into the developing mouse neocortex resulted in an increase in the proportion of radial glia undergoing cell division. Combined, these results suggest that, in normal conditions, PDGFD–PDGFR β signalling drives the proliferation of radial glia in humans, but not in mice (Fig. 1).

Nearly all murine radial glia reside in an area of the developing brain called the ventricular zone. In some species, however, additional progenitor populations take up residence adjacent to the ventricular zone, in another proliferative region called the subventricular zone (SVZ)^{3–7}. These cell types, including outer radial glia, have been implicated in the expansion of the neocortex of humans and other species^{3,4,7}. Lui and co-workers found that, in mice expressing the activated form of PDGFR β , cells resembling radial glia were distributed not only throughout the ventricular zone, but also throughout the SVZ. PDGFD-induced signalling may thus help to explain the establishment and proliferation of SVZ progenitors and, therefore, the expansion of the neocortex that has occurred in some mammals.

What particular type of progenitor cell Lui *et al.* observed in the SVZ remains unclear. Although these cells expressed a molecular marker for radial glia, they were not clearly identified as any one specific progenitor-cell population. Whether the effects of PDGFD–PDGFR β signalling extend to all types of neural progenitor is also yet to be investigated. Furthermore, because the authors compared radial glia only between human and mouse, we do not know whether the role of PDGFD is unique to humans or common to the development of radial glia in other species with an expanded neocortex. Finally, possible

associations between mutations in genes encoding members of the PDGF pathway and human neurodevelopmental abnormalities, in particular microcephaly (an abnormally small head circumference)^{2,10,11}, should now be pursued.

Perhaps most intriguing is the relatively limited number of genes whose expression seems to differ between human and murine developing radial glia. The authors developed a method to identify genes expressed in human radial glia and, as they had hoped, subsequently found a molecular pathway not shared by humans and mice. But the majority of the gene-expression data collected in this study emphasizes the remarkable commonalities between two transcriptional programs that diverged roughly 75 million years ago¹⁵. ■

Forrest O. Gulden and Nenad Šestan are in the Departments of Neurobiology and Psychiatry, Kavli Institute for Neuroscience and Program in Cellular Neuroscience, Neurodegeneration and Repair, Yale School of Medicine, New Haven, Connecticut 06510, USA.

NEUROSCIENCE

Towards unified vesicle endocytosis

An ultrafast, temperature-dependent mode of endocytosis, a process that is required for neurons to repeatedly fire, challenges current thinking and brings an old model back into the spotlight. [SEE ARTICLE P.228](#)

VLADAN LUČIĆ

Information is propagated between neurons at specialized junctions called synapses. Electrical signals generated at the presynaptic neuron cause neurotransmitter molecules to be released from synaptic vesicles into the synapse, and the neurotransmitter then binds to a receptor on the postsynaptic neuron, passing the signal on. To prevent depletion of synaptic vesicles during neuronal stimulation, vesicles are recycled through a process called endocytosis. Different types of synaptic-vesicle endocytosis have been described, but the mechanisms, and even the existence of one of the types, are still contested¹. In this issue, Watanabe *et al.*² (page 228) describe an ultrafast mode of endocytosis that operates at physiological temperatures. In doing so, they might provide a way to reconcile the diverging evidence on synaptic-vesicle endocytosis.

There are thought to be three main modes of synaptic-vesicle endocytosis, differing with respect to speed, capacity and the molecular

e-mails: forrest.gulden@yale.edu; nenad.sestan@yale.edu

1. Herculano-Houzel, S., Manger, P. R. & Kaas, J. H. *Front. Neuroanat.* **8**, 77 (2014).
2. Hill, R. S. & Walsh, C. A. *Nature* **437**, 64–67 (2005).
3. Dehay, C. & Kennedy, H. *Nature Rev. Neurosci.* **8**, 438–450 (2007).
4. Lui, J. H., Hansen, D. V. & Kriegstein, A. R. *Cell* **146**, 18–36 (2011).
5. Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S. & Kriegstein, A. R. *Nature* **409**, 714–720 (2001).
6. Breunig, J. J., Haydar, T. F. & Rakic, P. *Neuron* **70**, 614–625 (2011).
7. Taverna, E., Götz, M. & Huttner, W. B. *Annu. Rev. Cell Dev. Biol.* **30**, 465–502 (2014).
8. Lui, J. H. *et al.* *Nature* **515**, 264–268 (2014).
9. Evsyukova, I., Plestant, C. & Anton, E. S. *Annu. Rev. Cell Dev. Biol.* **29**, 299–353 (2013).
10. Thornton, G. K. & Woods, C. G. *Trends Genet.* **25**, 501–510 (2009).
11. Dixon-Salazar, T. J. & Gleeson, J. G. *Ann. NY Acad. Sci.* **1214**, 156–167 (2010).
12. Oldham, M. C. *et al.* *Nature Neurosci.* **11**, 1271–1282 (2008).
13. Kang, H. J. *et al.* *Nature* **478**, 483–489 (2011).
14. Bergsten, E. *et al.* *Nature Cell Biol.* **3**, 512–516 (2001).
15. Mouse Genome Sequencing Consortium *et al.* *Nature* **420**, 520–562 (2002).

machinery that regulates them¹ (Fig. 1a). So far, the best explained, and apparently predominant, mode is the clathrin-dependent pathway, in which synaptic vesicles merge with the cell's plasma membrane, and vesicles are regenerated directly from membrane-derived buds coated in clathrin protein. In the 'kiss-and-run' pathway, synaptic vesicles release a neurotransmitter through a transiently opened membrane pore and then quickly detach from the membrane without ever fully combining with it. Finally, in bulk endocytosis, which has the highest capacity of the three modes, large sections of the plasma membrane become internalized, forming endosomes from which synaptic vesicles can then bud.

Watanabe and colleagues combined sophisticated genetic techniques that stimulate neurons with high-pressure freezing to arrest endocytic processes at well-defined time points. They then determined the sequence of endocytic events by studying the process using electron microscopy, and showed that, at physiological temperatures (34–37°C),

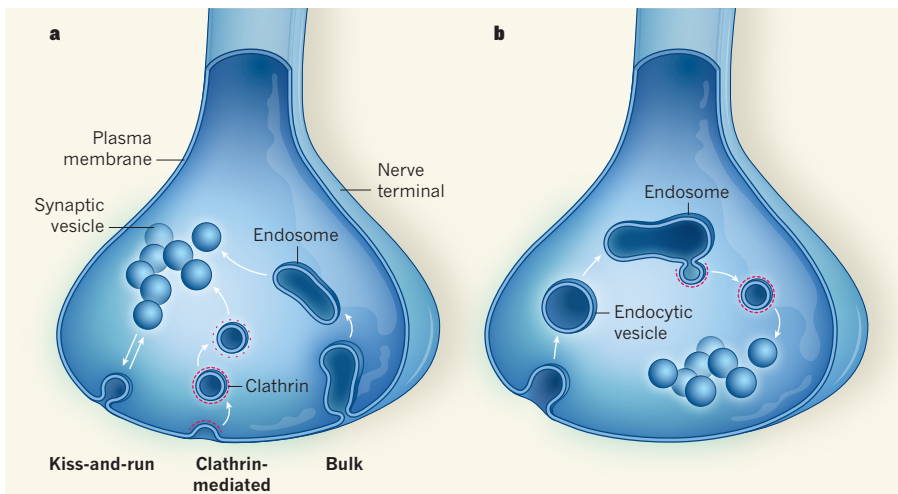


Figure 1 | Amalgamating endocytosis. **a**, The current models of synaptic-vesicle endocytosis. In kiss-and-run, synaptic vesicles form a transient pore in the cell's plasma membrane, and re-form as they detach and move into the cell. In clathrin-mediated endocytosis, synaptic vesicles are generated directly from clathrin-coated buds that form on the membrane. In bulk endocytosis, large sections of the membrane are internalized to form intracellular structures called endosomes, from which synaptic vesicles bud. **b**, Watanabe *et al.*² describe a mode of ultrafast endocytosis. Endosomes form from endocytic vesicles that arise as the membrane is internalized. Synaptic vesicles coated in clathrin then bud off from the endosomes within a few seconds of stimulation.

synaptic-vesicle endocytosis occurs very quickly. First, large endocytic vesicles bud inwards from the plasma membrane, then endosomes form, and finally synaptic vesicles bud off from them — all within a few seconds of neuronal stimulation (Fig. 1b).

The authors observed that genetic depletion of clathrin blocked endocytosis after endosome formation, implying that formation of endosomes, but not of synaptic vesicles, is clathrin independent. A similar endosome-dependent endocytic pathway was suggested 30 years ago³, in one of several landmark papers that established the role of synaptic vesicles in synaptic transmission, but this was dismissed in favour of the clathrin-dependent pathway. The arguments for clathrin-dependent endocytosis included the fact that many synaptic-vesicle precursors, in the form of clathrin-coated buds, were observed on the plasma membrane; that endosome formation was not seen in response to mild neuronal stimulation; and that clathrin-coated buds, but not endosomes, formed on protein-rich membrane regions. The technical advances of the past three decades allowed Watanabe and co-workers to show that clathrin-coated buds are located on endosomes, and that endosomes induced by mild neuronal stimulation are too short-lived to be detected by standard electron microscopy. Furthermore, they observed that clathrin-coated pits appear on the plasma membrane at room temperature (around 22 °C, the temperature at which the previous study³ was performed), but not at 34 °C.

Clathrin-dependent processes are thought to be much slower than the authors' model of endocytosis¹. However, in the past, most measurements have been taken at room

temperature, and biological processes (including synaptic transmission) are influenced by temperature — as a rule of thumb, there is a two- to threefold rate increase when the temperature increases by 10 °C. For instance, the role of synapsin protein in synaptic-vesicle organization and mobility is different at room and physiological temperatures⁴. Moreover, endocytosis occurs much faster at physiological temperatures than at room temperatures^{5,6}, with one paper⁷ reporting a rate comparable to that documented by Watanabe and colleagues.

How does temperature affect endocytic processes? The authors provide evidence to suggest that polymerization of the actin protein, which occurs at physiological but not at room temperature, is required for ultrafast endocytosis. There are diverging views on the role of actin in endocytosis that might be clarified by investigations at physiological temperatures. Rates of protein phosphorylation and dephosphorylation are also candidate effectors, because they are temperature-dependent and implicated in synaptic-vesicle endocytosis.

Endosome formation followed by synaptic-vesicle formation from endosomes occurs in both bulk endocytosis and Watanabe and co-workers' ultrafast pathway. In addition, evidence suggests that clathrin is involved in synaptic-vesicle, but not endosome, formation during bulk endocytosis⁷. At room temperature, bulk endocytosis is induced by strong neuronal stimulation⁸, inviting speculation that the bulk and ultrafast pathways are two faces of the same process, in which either strong stimulation or increased temperature can induce rapid endosome formation.

Similarities between the ultrafast and the other two endocytic pathways also exist.

A significant body of literature is devoted to the steps by which clathrin-coated pits are formed on the membrane and synaptic vesicles are generated from these pits. If endosomes are formed by rapid internalization of the plasma membrane, their lipid composition and protein content may be the same as that of the plasma membrane. This would suggest that synaptic-vesicle formation from endosomes follows the same biochemical pathways as those described for the clathrin-dependent process. Finally, the sub-second timescale of membrane internalization in ultrafast endocytosis mimics arguably the most striking feature of the kiss-and-run pathway — its speed⁹.

There are many molecular details to clarify before a common basis for endocytosis can be established. For example, two papers published this year^{10,11} point to different roles for clathrin and related proteins in bulk endocytosis. Different stimulation conditions and detection methods might have led to these disparate conclusions. This should be considered in follow-up work to confirm that the stimulation conditions and genetic manipulations used in the current paper do not affect endocytosis.

In summary, Watanabe and co-workers find temperature to be a major determinant of endocytosis. The ultrafast endocytosis they describe combines aspects of other endocytic pathways — it depends on clathrin for synaptic-vesicle formation, reaches the speed of kiss-and-run and requires formation of endosomes from membrane internalization. A common pathway for endocytosis would explain how a well-defined structure such as the synaptic vesicle is seemingly generated by different mechanisms¹. It is stimulating to speculate that the three previously defined modes of endocytosis can in fact be unified into one, which can vary under different conditions. However, it is too early to say for sure. Hopefully, this and other recent findings^{10,11} will provoke more experiments at physiological temperatures, and a careful reassessment of the existing literature. ■

Vladan Lučić is at the Max Planck Institute of Biochemistry, 82152 Martinsried, Germany.
e-mail: vladan@biochem.mpg.de

1. Saheki, Y. & De Camilli, P. *Cold Spring Harb. Perspect. Biol.* **4**, a005645 (2012).
2. Watanabe, S. *et al.* *Nature* **515**, 228–233 (2014).
3. Miller, T. M. & Heuser, J. E. *J. Cell Biol.* **98**, 685–698 (1984).
4. Gaffield, M. A. & Betz, W. J. *J. Neurosci.* **27**, 13691–13700 (2007).
5. Micheva, K. D. & Smith, S. J. *J. Neurosci.* **25**, 7481–7488 (2005).
6. Renden, R. & von Gersdorff, H. *J. Neurophysiol.* **98**, 3349–3359 (2007).
7. Armbruster, M., Messa, M., Ferguson, S. M., De Camilli, P. & Ryan, T. A. *eLife* **2**, e00845 (2013).
8. Clayton, E. L. & Cousin, M. A. *J. Neurochem.* **111**, 901–914 (2009).
9. Rizzoli, S. O. & Jahn, R. *Traffic* **8**, 1137–1144 (2007).
10. Kononenko, N. L. *et al.* *Neuron* **82**, 981–988 (2014).
11. Wu, Y. *et al.* *eLife* **3**, e01621 (2014).

This article was published online on 8 October 2014.

Synaptic, transcriptional and chromatin genes disrupted in autism

A list of authors and their affiliations appears at the end of the paper

The genetic architecture of autism spectrum disorder involves the interplay of common and rare variants and their impact on hundreds of genes. Using exome sequencing, here we show that analysis of rare coding variation in 3,871 autism cases and 9,937 ancestry-matched or parental controls implicates 22 autosomal genes at a false discovery rate (FDR) < 0.05, plus a set of 107 autosomal genes strongly enriched for those likely to affect risk (FDR < 0.30). These 107 genes, which show unusual evolutionary constraint against mutations, incur *de novo* loss-of-function mutations in over 5% of autistic subjects. Many of the genes implicated encode proteins for synaptic formation, transcriptional regulation and chromatin-remodelling pathways. These include voltage-gated ion channels regulating the propagation of action potentials, pacemaking and excitability-transcription coupling, as well as histone-modifying enzymes and chromatin remodellers—most prominently those that mediate post-translational lysine methylation/demethylation modifications of histones.

Features of subjects with autism spectrum disorder (ASD) include compromised social communication and interaction. Because the bulk of risk arises from *de novo* and inherited genetic variation^{1–10}, characterizing which genes are involved informs ASD neurobiology and reveals part of what makes us social beings.

Whole-exome sequencing (WES) studies have proved fruitful in uncovering risk-conferring variation, especially by enumerating *de novo* variation, which is sufficiently rare that recurrent mutations in a gene provide strong evidence for a causal link to ASD. *De novo* loss-of-function (LoF) single-nucleotide variants (SNVs) or insertion/deletion (indel) variants^{11–15} are found in 6.7% more ASD subjects than in matched controls and implicate nine genes from the first 1,000 ASD subjects analysed^{11–16}. Moreover, because there are hundreds of genes involved in ASD risk, ongoing WES studies should identify additional ASD genes as an almost linear function of increasing sample size¹¹.

Here we conduct the largest ASD WES study so far, analysing 16 sample sets comprising 15,480 DNA samples (Supplementary Table 1 and Extended Data Fig. 1). Unlike earlier WES studies, we do not rely solely on counting *de novo* LoF variants, rather we use novel statistical methods to assess association for autosomal genes by integrating *de novo*, inherited and case-control LoF counts, as well as *de novo* missense variants predicted to be damaging. For many samples original data from sequencing performed on Illumina HiSeq 2000 systems were used to call SNVs and indels in a single large batch using GATK (v2.6)¹⁷. *De novo* mutations were called using enhancements of earlier methods¹⁴ (Supplementary Information), with calls validating at extremely high rates.

After evaluation of data quality, high-quality alternative alleles with a frequency of <0.1% were identified, restricted to LoF (frameshifts, stop gains, donor/acceptor splice site mutations) or probably damaging missense (Mis3) variants (defined by PolyPhen-2 (ref. 18)). Variants were classified by type (*de novo*, case, control, transmitted, non-transmitted) and severity (LoF, Mis3), and counts tallied for each gene.

Some 13.8% of the 2,270 ASD trios (two parents and one affected child) carried a *de novo* LoF mutation—significantly in excess of both the expected value¹⁹ (8.6%, $P < 10^{-14}$) and what was observed in 510 control trios (7.1%, $P = 1.6 \times 10^{-5}$) collected here and previously published¹⁵. Eighteen genes (Table 1) exhibited two or more *de novo* LoF mutations. These genes are all known or strong candidate ASD genes, but given the number of trios sequenced and gene mutability^{14,19}, we

would expect to observe this in approximately two such genes by chance. While we expect only two *de novo* Mis3 events in these 18 genes, we observe 16 ($P = 9.2 \times 10^{-11}$, Poisson test). Because most of our data exist in cases and controls and because we observed an additional excess of transmitted LoF events in the 18 genes, it is evident that the optimal analytical framework must involve an integration of *de novo* mutation with variants observed in cases and controls and transmitted or untransmitted from carrier parents. Investigating beyond *de novo* LoFs is also critical given that many ASD risk genes and loci have mutations that are not completely penetrant.

Transmission and *de novo* association

We adopted TADA (transmission and *de novo* association), a weighted, statistical model integrating *de novo*, transmitted and case-control variation²⁰. TADA uses a Bayesian gene-based likelihood model including per-gene mutation rates, allele frequencies, and relative risks of particular classes of sequence changes. We modelled both LoF and Mis3 sequence variants. Because no aggregate association signal was detected for inherited Mis3 variants, they were not included in the analysis. For each gene, variants of each class were assigned the same effect on relative risk. Using a prior probability distribution of relative risk across genes for each class of variants, the model effectively weighted different classes of variants in this order: *de novo* LoF > *de novo* Mis3 > transmitted LoF, and allowed for a distribution of relative risks across genes for each class. The strength of association was assimilated across classes to produce a gene-level Bayes factor with a corresponding FDR q value. This framework increases the power compared to the use of *de novo* LoF variants alone (Extended Data Fig. 2).

TADA identified 33 autosomal genes with an FDR < 0.1 (Table 1) and 107 with an FDR < 0.3 (Supplementary Tables 2 and 3 and Extended Data Fig. 3). Of the 33 genes, 15 (45.5%) are known ASD risk genes⁹; 11 have been reported previously with mutations in ASD patients but were not classed as true risk genes owing to insufficient evidence (*SUV420H1* (refs 11, 15), *ADNP*¹², *BCL11A*¹⁵, *CACNA2D3* (refs 15, 21), *CTTNBP2* (ref. 15), *GABRB3* (ref. 21), *CDC42BPB*¹³, *APH1A*¹⁴, *NR3C2* (ref. 15), *SETD5* (refs 14, 22) and *TRIO*¹¹) and 7 are completely novel (*ASHIL*, *MLL3* (also known as *KMT2C*), *ETFB*, *NAA15*, *MYO9B*, *MIB1* and *VIL1*). *ADNP* mutations have recently been identified in 10 patients with ASD and other shared clinical features²³. Two of the newly discovered genes,

Table 1 | ASD risk genes

dnLoF count	FDR ≤ 0.01	0.01 < FDR ≤ 0.05	0.05 < FDR ≤ 0.1
≥2	<i>ADNP</i> , <i>ANK2</i> , <i>ARID1B</i> , <i>CHD8</i> , <i>CUL3</i> , <i>DYRK1A</i> , <i>GRIN2B</i> , <i>KATNAL2</i> , <i>POGZ</i> , <i>SCN2A</i> , <i>SUV420H1</i> , <i>SYNGAP1</i> , <i>TBR1</i>	<i>ASXL3</i> , <i>BCL11A</i> , <i>CACNA2D3</i> , <i>MLL3</i>	<i>ASH1L</i>
1		<i>CTTNBP2</i> , <i>GABRB3</i> , <i>PTEN</i> , <i>RELN</i>	<i>APH1A</i> , <i>CD42BPB</i> , <i>ETFB</i> , <i>NAA15</i> , <i>MYO9B</i> , <i>MYT1L</i> , <i>NR3C2</i> , <i>SETD5</i> , <i>TRIO</i>
0		<i>MIB1</i>	<i>VIL1</i>

TADA analysis of LoF and damaging missense variants found to be *de novo* in ASD subjects, inherited by ASD subjects, or present in ASD subjects (versus control subjects). dnLoF, *de novo* LoF events.

ASH1L and *MLL3*, converge on chromatin remodelling. *MYO9B* plays a key role in dendritic arborization²⁴. *MIB1* encodes an E3 ubiquitin ligase critical for neurogenesis²⁵ and is regulated by miR-137 (ref. 26), a microRNA that regulates neuronal maturation and is implicated in schizophrenia risk²⁷.

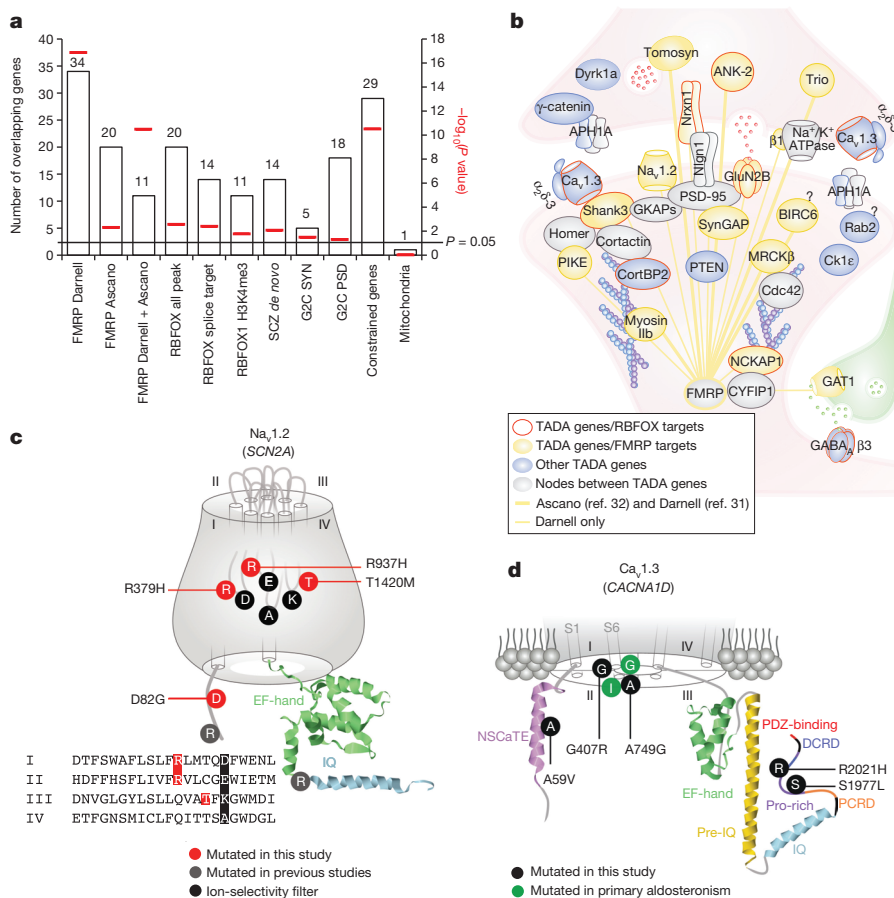
When the WES data from genes with an FDR < 0.3 were evaluated for the presence of deletion copy number variants (CNVs) (such CNVs are functionally equivalent to LoF mutations), 34 CNVs meeting quality and frequency constraints (Supplementary Information) were detected in 5,781 samples (Extended Data Fig. 1). Of the 33 genes with an FDR < 0.1, 3 contained deletion CNVs mapping to 3 ASD subjects and one parent. Of the 74 genes meeting the criterion $0.1 \leq \text{FDR} < 0.3$, about one-third could be false positives. Deletion CNVs were found in 14 of these genes and the data supported risk status for 10 of them (Extended Data Table 1 and Extended Data Fig. 4). Two of these ten, *NRXN1* and *SHANK3*, were previously implicated in ASD^{2,3,10}. The risk from deletion CNVs, as measured by the odds ratio, is comparable to that from LoF SNVs in cases versus controls or transmission of LoF variants from parents to offspring.

Estimated odds ratios of top genes

Inherent in our conception of the biology of ASD is the notion that there is variation between genes in their impact on risk; for a given

class of variants (for example, LoF) some genes have a large impact, others smaller, and still others have no effect at all. In addition, mis-annotation of variants, among other confounds, can yield false variant calls in subjects (Supplementary Information). These confounds can often be overcome by examining the data in a manner orthogonal to gene discovery. For example, females have greatly reduced rates of ASD relative to males (a 'female protective effect'). Consequently, and regardless of whether this is diagnostic bias or biological protection, females have a higher liability threshold, requiring a larger genetic burden before being diagnosed^{22,28,29}. A corollary is that if a variant has the same effect on autism liability in males as it does in females, that variant will be present at a higher frequency in female ASD cases compared to males. Importantly, the magnitude of the difference is proportional to risk as measured by the odds ratio; hence, the effect on risk for a class of variants can be estimated from the difference in frequency between males and females.

Genes with an FDR < 0.1 show profound female enrichment for *de novo* events ($P = 0.005$ for LoF, $P = 0.004$ for Mis3), consistent with *de novo* events having large impacts on liability (odds ratio ≥ 20; Extended Data Fig. 5). However, genes with an FDR between 0.1 and 0.3 show substantially less enrichment for female events, consistent with a modest impact for LoF variants (odds ratio range 2–4, whether transmitted or *de novo*) and little to no effect from Mis3 variants. The



results are consistent with inheritance patterns: LoF mutations in $FDR < 0.1$ genes are rarely inherited from unaffected parents whereas those in the $0.1 \leq FDR < 0.3$ group are far more often inherited than they are *de novo* mutations.

By analysing the distribution of relative risk over inferred ASD genes²⁰, the number of ASD risk genes can be estimated. The estimate relies on the balance of genes with multiple *de novo* LoF mutations versus those with only one: the larger the number of ASD genes, the greater proportion that will show only one *de novo* LoF. This approach yields an estimate of 1,150 ASD genes (Supplementary Information). While there are many more genes to be discovered, many will have a modest impact on risk compared to the genes in Table 1.

Enrichment analyses

Gene sets with an $FDR < 0.3$ are strongly enriched for genes under evolutionary constraint¹⁹ ($P = 3.0 \times 10^{-11}$; Fig. 1a and Supplementary Table 4), consistent with the hypothesis that heterozygous LoF mutations in these genes are ASD risk factors. Over 5% of ASD subjects carry *de novo* LoF mutations in our $FDR < 0.3$ list. We also observed that genes in the $FDR < 0.3$ list had a significant excess of *de novo* non-synonymous events detected by the largest schizophrenia WES study so far³⁰ ($P = 0.0085$; Fig. 1a), providing further evidence for overlapping risk loci between these disorders and independent confirmation of the signal in the gene sets presented here.

We found significant enrichment for genes encoding messenger RNAs targeted by two neuronal RNA-binding proteins: FMRP³¹ (also known as FMR1), mutated or absent in fragile X syndrome ($P = 1.20 \times 10^{-17}$, 34 targets³¹, of which 11 are corroborated by an independent data set³²), and RBFOX (RBFOX1/2/3) ($P = 0.0024$, 20 targets, of which 12 overlap with FMRP), with RBFOX1 shown to be a splicing factor dysregulated in ASD^{33,34} (Fig. 1a). These two pathways expand the complexity of ASD neurobiology to post-transcriptional events, including splicing and translation, both of which sculpt the neural proteome.

We found nominal enrichment for human orthologues of mouse genes encoding synaptic ($P = 0.031$) and post-synaptic density (PSD) proteins³⁵ ($P = 0.046$; Fig. 1a, b and Supplementary Tables 4–6). Enrichment analyses for InterPro, SMART or Pfam domains ($FDR < 0.05$ and a minimum of five genes per category) reveal an overrepresentation of DNA- or histone-related domains: eight genes encoding proteins with InterPro zinc-finger FYVE PHD domains (142 such annotated genes in the genome; $FDR = 7.6 \times 10^{-4}$), and five with Pfam Su(var)3-9, enhancer-of-zeste, trithorax (SET) domains (39 annotated in the genome; $FDR = 8.2 \times 10^{-4}$).

Integrating complementary data

To implicate additional genes in risk for ASD, we used a model called DAWN (detecting association with networks)³⁶. DAWN evokes a hidden Markov random field framework to identify clusters of genes that show strong association signals and highly correlated co-expression in a key tissue and developmental context. Previous research suggests human mid-fetal prefrontal and motor-somatosensory neocortex is a critical nexus for risk¹⁶, thus we evaluated gene co-expression data from that tissue together with TADA scores for genes with an $FDR < 0.3$. Because this list is enriched for genes under evolutionary constraint, we generalized DAWN to incorporate constraint scores (Supplementary Information). When TADA results, gene co-expression in mid-fetal neocortex and constraint scores are jointly modelled, DAWN identifies 160 genes that plausibly affect risk (Fig. 2), 91 of which are not in the 107 TADA genes with an $FDR < 0.3$. Moreover, the model parameter describing evolutionary constraint is an important predictor of clusters of putative risk genes ($P = 0.018$).

A subnetwork obtained by seeding the 160 DAWN genes within a high-confidence protein–protein interactome¹⁴ confirmed that the putative genes are enriched for neuronal functions. We kept the largest connected component, containing 95 seed DAWN genes, 50 of which were in the $FDR < 0.3$ gene set. The DAWN gene products form four natural

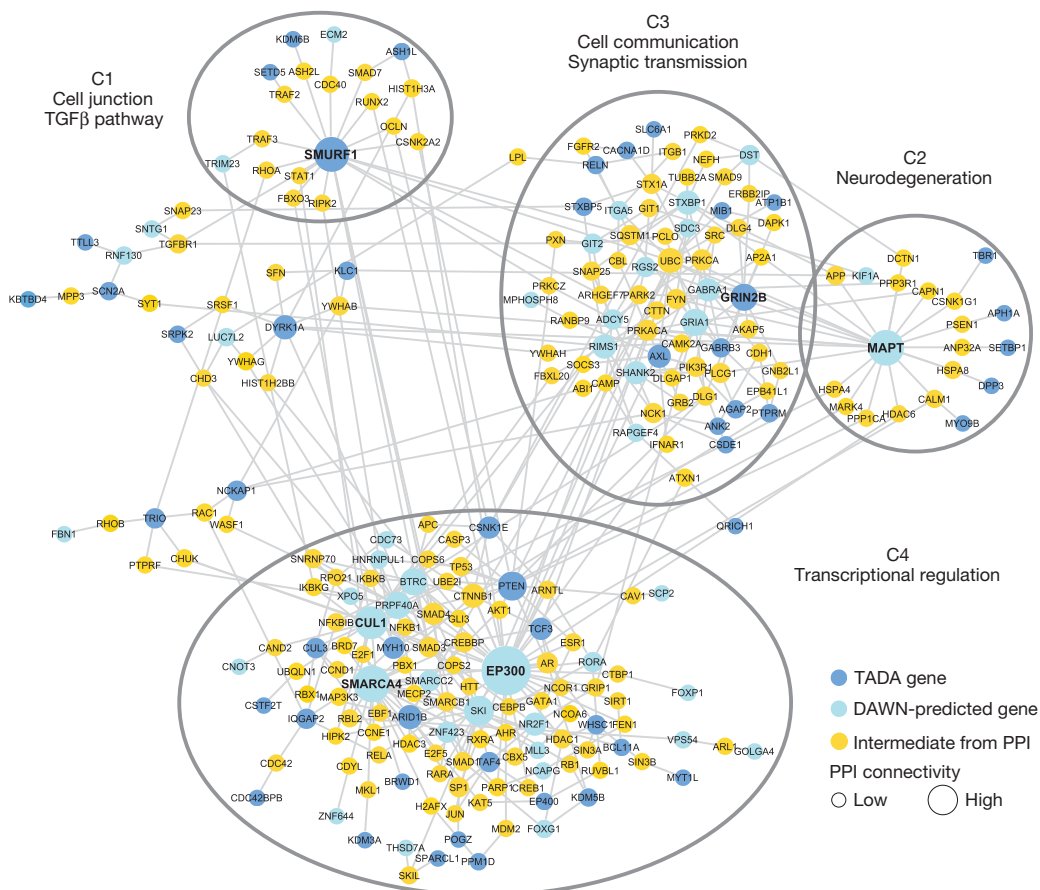


Figure 2 | ASD genes in neuronal networks. Protein–protein interaction network created by seeding TADA and DAWN-predicted genes. Only intermediate genes that are known to interact with at least two TADA and/or DAWN genes are included. Four natural clusters (C1–C4) are demarcated with black ellipses. All nodes are sized on the basis of degree of connectivity.

clusters on the basis of network connectivity (Fig. 2). We visualized the enriched pathways and biological functions for each of these clusters on ‘cavases’³⁷ (Extended Data Fig. 6). Many of the previously known ASD risk genes fall in cluster C3, including genes involved in synaptic transmission and cell–cell communication. Cluster C4 is enriched for genes related to transcriptional and chromatin regulation. Many TADA and DAWN genes in this cluster interact tightly with other transcription factors, histone-modifying enzymes and DNA-binding proteins. Five TADA genes in the cluster C2 are bridged to the rest of the network through *MAPT*, as inferred by DAWN. The enrichment results for cluster C2 indicate that genes implicated in neurodegenerative disorders could also have a role in neurodevelopmental disorders.

Emergent results

Amongst the critical synaptic components found to be mutated in our study are voltage-gated ion channels involved in fundamental processes including the propagation of action potentials (for example, the $\text{Na}_v1.2$ channel), neuronal pacemaking and excitability–transcription coupling (for example, the $\text{Ca}_v1.3$ channel) (Fig. 1b). We identified four LoF and five Mis3 variants in *SCN2A* ($\text{Na}_v1.2$), three Mis3 variants in *CACNA1D* ($\text{Ca}_v1.3$) and two LoF variants in *CACNA2D3* ($\alpha_2\delta$ -3 subunit). Remarkably, three *de novo* Mis3 variants in *SCN2A* affected residues mutated in homologous genes in patients with other syndromes, including Brugada syndrome (*SCN5A*) or epilepsy disorders (*SCN1A*) (Arg379His and Arg937His). These arginines, as well as the threonine mutated in Thr1420Met, cluster to the P-loops forming the ion selectivity filter, located in proximity to the inner ring (DEKA motif) (Fig. 1c). Because homologous channels mutated in these arginines do not conduct inward Na^+ currents^{38,39}, Arg379His and Arg937His mutations might have similar effect.

Two *de novo* *CACNA1D* variants (Gly407Arg and Ala749Gly) emerged at positions proximal to residues mutated in patients with primary aldosteronism and neurological deficits (Fig. 1d). The reported mutations interfere with channel activation and inactivation⁴⁰. Amongst variants found in cases, Ala59Val maps to the NSCaTE domain, also important for Ca^{2+} -dependent inactivation, and Ser1977Leu and Arg2021His co-cluster in the carboxy-terminal proline-rich domain, the site of interaction with SHANK3, a key PSD scaffolding protein. Mutations in *RIMS1* and *RIMBP2*, which can associate with $\text{Ca}_v1.3$, were found in our cohort (but with an FDR > 0.3).

Chromatin remodelling involves histone-modifying enzymes (encoded by histone-modifier genes, HMGs) and chromatin remodellers (readers) that recognize specific histone post-translational modifications and orchestrate their effects on chromatin. Our gene set is enriched in HMGs (9 HMGs out of 152 annotated in Histome⁴¹, Fisher’s exact test, $P = 2.2 \times 10^{-7}$). Enrichment in the gene ontology term ‘histone-lysine N-methyltransferase activity’ (5 genes out of 41 so annotated; FDR = 2.2×10^{-2}) highlights this as a prominent pathway.

Lysines on histones 3 and 4 can be mono-, di- or tri-methylated, providing a versatile mechanism for either activation or repression of transcription. Of 107 TADA genes, five are SET lysine methyltransferases, four are jumoni lysine demethylases, and two are readers (Fig. 3a). RBFOX1 co-isolates with histone H3 trimethyl Lys 4 (H3K4me3)⁴², and our data set is enriched in targets shared by RBFOX1 and H3K4me3 ($P = 0.0166$; Fig. 1a and Supplementary Table 4). Some *de novo* missense variants targeting these genes map to functional domains (Extended Data Fig. 7).

For the H3K4me2 reader *CHD8*, we extended our analyses in search of additional *de novo* variation in the cases of the case-control sample. By sequencing complete parent–child trios for many *CHD8* variants, five variants were found to be *de novo*, two of which affect essential splice sites and cause LoF by exon skipping or activation of cryptic splice sites in lymphoblastoid cells (Fig. 3b).

Given the role of HMGs in transcription, we reasoned that TADA genes might be interconnected through transcription ‘routes’. We searched for a connected network (seeded by 9 TADA HMGs) in a transcription factor interaction network (ChEA)⁴³. We found that 46 TADA genes

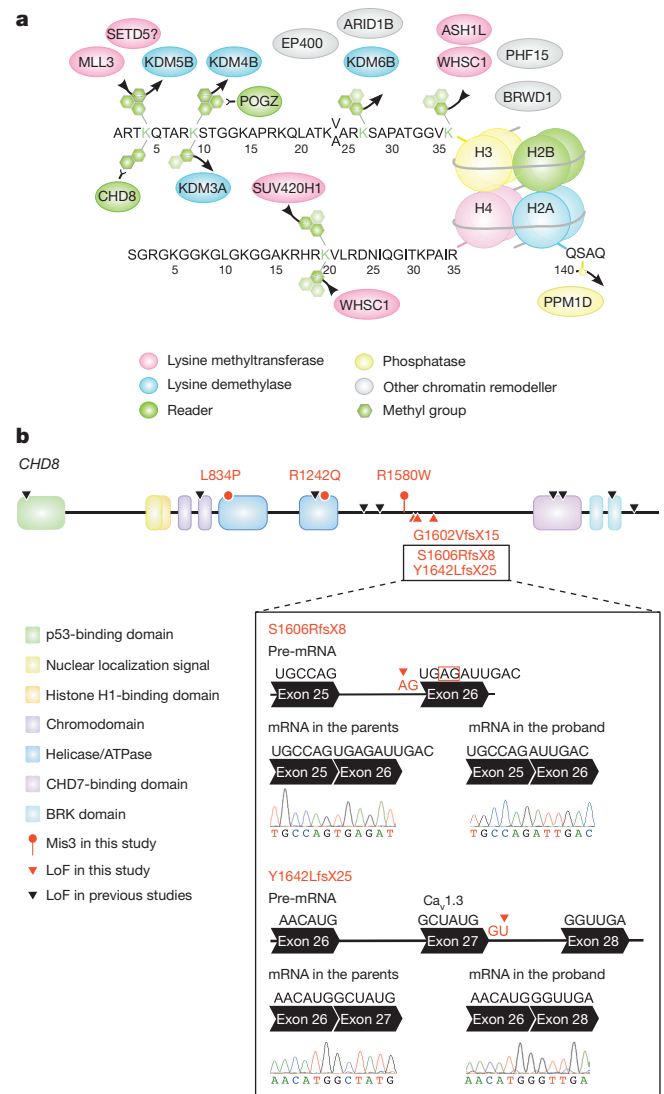


Figure 3 | ASD genes in chromatin remodelling. **a**, TADA genes cluster to chromatin-remodelling complexes. Amino-terminals of histones H3, H4 and part of H2A are shown. Lysine methyltransferases add methyl groups, whereas lysine demethylases remove them. **b**, *De novo* Mis3 and LoF variants in *CHD8*. The box shows the outcome of reverse transcription PCR and Sanger sequencing in lymphoblastoid cells for two newly identified *de novo* splice-site variants. The first mutation affects an acceptor splice site (red arrow), causing the activation of a cryptic splice site (red box), a four-nucleotide deletion, frame shift and a premature stop. The second mutation affects a donor splice site (red arrow), causing exon skipping, frame shift and a premature stop.

are directly interconnected in a 55-gene cluster (Extended Data Fig. 8) ($P = 0.002$; 1,000 random draws), for a total of 69 when including all known HMGs (Fig. 4) ($P = 0.001$; 1,000 random draws).

Examining the Human Gene Mutation Database we found that the 107 TADA genes included 21 candidate genes for intellectual disability, 3 for epilepsy, 17 for schizophrenia, 9 for congenital heart disease and 6 for metabolic disorders (Fig. 5).

Conclusions

Complementing earlier reports, ASD subjects show a clear excess of *de novo* LoF mutations above expectation, with a concentration of such events in a handful of genes. While this handful has a large effect on risk, most ASD genes have a much smaller impact. This gradient emerges most notably from the contrast of risk variation in male and female ASD subjects. Unlike some earlier studies, but consistent with expectation, the data also show clear evidence for effect of *de novo* missense SNVs

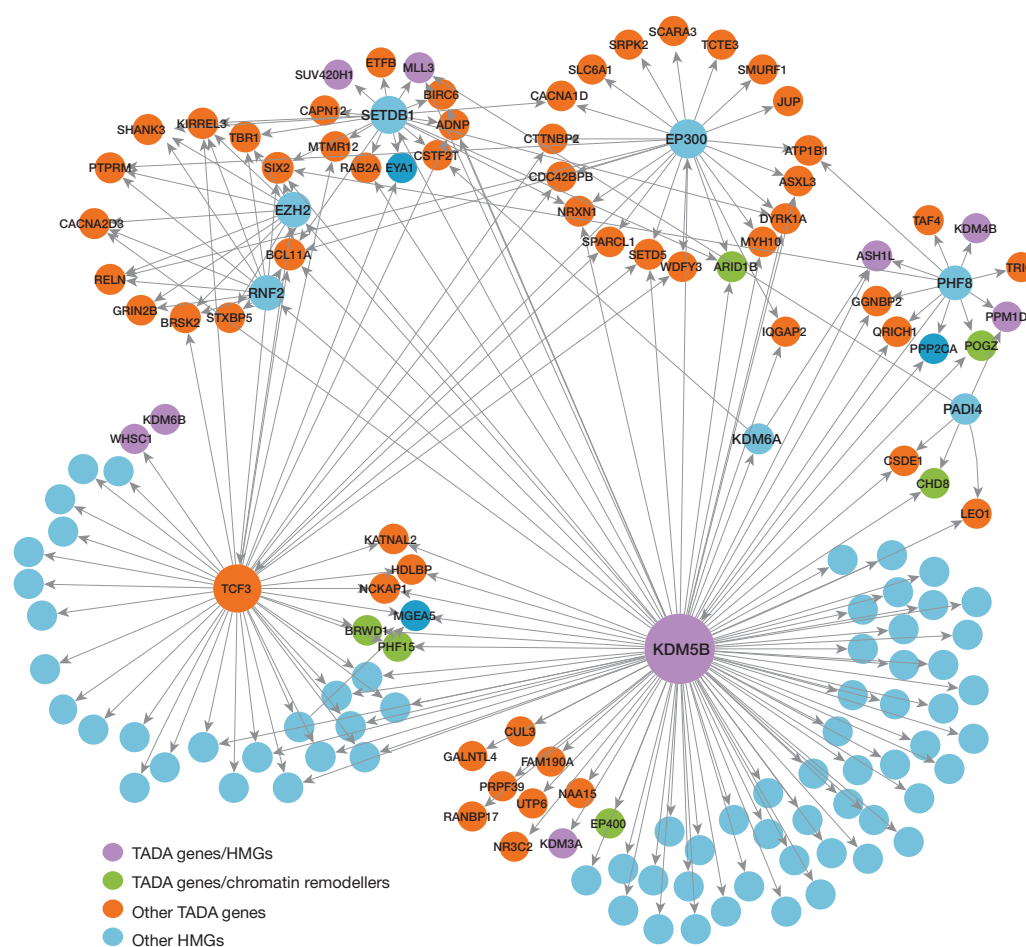


Figure 4 | Transcription regulation network of TADA genes. Edges indicate transcription regulators (source nodes) and their gene targets (target nodes) based on the ChEA network; interactions among only HMGs are ignored.

on risk; for risk generated by LoF variants transmitted from unaffected parents; and for the value of case-control design in gene discovery. By integrating data on *de novo*, inherited and case-control variation, the yield of ASD gene discoveries was doubled over what would be obtained from a count of *de novo* LoF variants alone. ASD genes almost uniformly show strong constraints against variation, a feature we exploit to implicate other genes in risk.

Three critical pathways for typical development are damaged by risk variation: chromatin remodelling, transcription and splicing, and synaptic function. Chromatin remodelling controls events underlying

the formation of neural connections, including neurogenesis and neural differentiation⁴⁴, and relies on epigenetic marks as post-translational modifications of histones. Here we provide extensive evidence for HMGs and readers in sporadic ASD, implicating specifically lysine methylation and extending the mutational landscape of the emergent ASD gene *CHD8* to missense variants. Splicing is implicated by the enrichment of RBFOX targets in the top ASD candidates. Risk variation also affects multiple classes and components of synaptic networks, from receptors and ion channels to scaffolding proteins. Because a wide set of synaptic genes is disrupted in idiopathic ASD, it seems reasonable to suggest that altered chromatin dynamics and transcription, induced by disruption of relevant genes, leads to impaired synaptic function as well. *De novo* mutations in ASD^{11–15}, intellectual disability⁴⁵ and schizophrenia³⁰ cluster to synaptic genes, and synaptic defects have been reported in models of these disorders⁴⁶. Integrity of synaptic function is essential for neural physiology, and its perturbation could represent the intersection between diverse neuropsychiatric disorders⁴⁷.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 May; accepted 18 August 2014.

Published online 29 October; corrected online 12 November 2014 (see full-text HTML version for details).

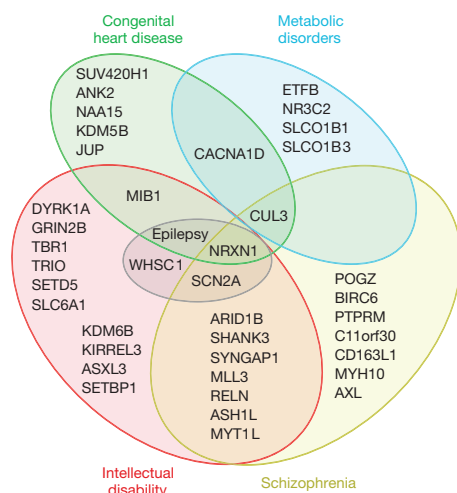


Figure 5 | Involvement in disease of ASD genes. The Venn diagram shows the overlap in disease involvement for the TADA genes.

1. Ronald, A. & Hoekstra, R. A. Autism spectrum disorders and autistic traits: a decade of new twin studies. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **156**, 255–274 (2011).
2. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
3. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
4. Klei, L. *et al.* Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism* **3**, 9 (2012).

5. Gaugler, T. *et al.* Most inherited risk for autism resides with common variation. *Nature Genet.* **46**, 881–885 (2014).
6. Yu, T. W. *et al.* Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77**, 259–273 (2013).
7. Lim, E. T. *et al.* Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* **77**, 235–242 (2013).
8. Poultnery, C. S. *et al.* Identification of small exonic CNV from whole-exome sequencing data and application to autism spectrum disorder. *Am. J. Hum. Genet.* **93**, 607–619 (2013).
9. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42–77 (2011).
10. Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569–573 (2009).
11. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
12. O’Roak, B. J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
13. O’Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246–250 (2012).
14. Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
15. Iossifov, I. *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
16. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
17. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).
18. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).
19. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nature Genet.* **46**, 944–950 (2014).
20. He, X. *et al.* Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet.* **9**, e1003671 (2013).
21. Girirajan, S. *et al.* Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am. J. Hum. Genet.* **92**, 221–237 (2013).
22. Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* **94**, 677–694 (2014).
23. Helsmoortel, C. *et al.* A SWI/SNF-related autism syndrome caused by de novo mutations in *ADNP*. *Nature Genet.* **46**, 380–384 (2014).
24. Long, H. *et al.* Myo9b and RICS modulate dendritic morphology of cortical neurons. *Cereb. Cortex* **23**, 71–79 (2013).
25. Yoon, K. J. *et al.* Mind bomb 1-expressing intermediate progenitors generate Notch signaling to maintain radial glial cells. *Neuron* **58**, 519–531 (2008).
26. Smrt, R. D. *et al.* MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase Mind bomb-1. *Stem Cells* **28**, 1060–1070 (2010).
27. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genet.* **45**, 1150–1159 (2013).
28. Robinson, E. B., Lichtenstein, P., Anckarsater, H., Hapke, F. & Ronald, A. Examining and interpreting the female protective effect against autistic behavior. *Proc. Natl Acad. Sci. USA* **110**, 5258–5262 (2013).
29. Jacquemont, S. *et al.* A higher mutational burden in females supports a “female protective model” in neurodevelopmental disorders. *Am. J. Hum. Genet.* **94**, 415–425 (2014).
30. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
31. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
32. Ascano, M. Jr. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492**, 382–386 (2012).
33. Weyn-Vanhenhenryck, S. M. *et al.* HITS-CLIP and integrative modeling define the Rbox splicing-regulatory network linked to brain development and autism. *Cell Rep.* **6**, 1139–1152 (2014).
34. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
35. Collins, M. O. *et al.* Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. *J. Neurochem.* **97** (suppl. 1), 16–23 (2006).
36. Liu, L. *et al.* DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. *Mol. Autism* **5**, 22 (2014).
37. Tan, C. M., Chen, E. Y., Dannenfelser, R., Clark, N. R. & Ma’ayan, A. Network2Canvas: network visualization on a canvas with enrichment analysis. *Bioinformatics* **29**, 1872–1878 (2013).
38. Vatta, M. *et al.* Genetic and biophysical basis of sudden unexplained nocturnal death syndrome (SUNDS), a disease allelic to Brugada syndrome. *Hum. Mol. Genet.* **11**, 337–345 (2002).
39. Volkler, L. *et al.* Na_v 1.1 dysfunction in genetic epilepsy with febrile seizures-plus or Dravet syndrome. *Eur. J. Neurosci.* **34**, 1268–1275 (2011).
40. Scholl, U. I. *et al.* Somatic and germline *CACNA1D* calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. *Nature Genet.* **45**, 1050–1054 (2013).
41. Khare, S. P. *et al.* Histone—a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res.* **40**, D337–D342 (2012).
42. Feng, J. *et al.* Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. *Genome Biol.* **15**, R65 (2014).
43. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
44. Ronan, J. L., Wu, W. & Crabtree, G. R. From neural development to cognition: unexpected roles for chromatin. *Nature Rev. Genet.* **14**, 347–359 (2013).
45. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
46. Penzes, P., Cahill, M. E., Jones, K. A., VanLeeuwen, J. E. & Woolfrey, K. M. Dendritic spine pathology in neuropsychiatric disorders. *Nature Neurosci.* **14**, 285–293 (2011).
47. Zoghbi, H. Y. Postnatal neurodevelopmental disorders: meeting at the synapse? *Science* **302**, 826–830 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by National Institutes of Health (NIH) grants U01MH100233, U01MH100209, U01MH100229 and U01MH100239 to the Autism Sequencing Consortium. Sequencing at Broad Institute was supported by NIH grants R01MH089208 (M.J.D.) and new sequencing by U54 HG003067 (S. Gabriel, E. Lander). Other funding includes NIH R01 MH089482, R37 MH057881 (B.D. and K.R.), R01 MH061009 (J.S.S.), UL1TR000445 (NCAT to VUMC); P50 HD055751 (E.H.C.); MH089482 (J.S.S.), NIH R01 MH083565 and RC2MH089952 (C.A.W.), NIMH MH095034 (P.S.), MH077139 (P.F. Sullivan); 5UL1 RR024975 and P30 HD15052. The DDD Study is funded by HICF-1009-003 and WT098051. UK10K is funded by WT091310. We also acknowledge The National Children’s Research Foundation, Our Lady’s Children’s Hospital, Crumlin; The Meath Foundation; AMNCH, Tallaght; The Health Research Board, Ireland and Autism Speaks, U.S.A. C.A.W. is an Investigator of the Howard Hughes Medical Institute. S.D.R., A.P.G., C.S.P., Y.K. and S.-C.F. are Seaver fellows, supported by the Seaver foundation. A.P.G. is also supported by the Charles and Ann Schlaifer Memorial Fund. P.F.B. is supported by a UK National Institute for Health Research (NIHR) Senior Investigator award and the NIHR Biomedical Research Centre in Mental Health at the South London & Maudsley Hospital. A.C. is supported by María José Jove Foundation and the grant FIS PI13/O1136 of the Strategic Action from Health Carlos III Institute (FEDER). This work was supported in part through the computational resources and staff expertise provided by the Department of Scientific Computing at the Icahn School of Medicine at Mount Sinai. We acknowledge the assistance of D. Hall and his team at National Database for Autism Research. We thank Jian Feng for providing a list of targets of both RBFOX1 and H3K4me3. We thank M. Potter for data coordination; K. Moore and J. Reichert for technical assistance; and, S. Lindsay for helping with molecular validation. We acknowledge the clinicians and organizations that contributed to samples used in this study. Finally, we are grateful to the many families whose participation made this study possible.

Author Contributions Study conception and design: J.D.B., D.J.C., M.J.D., S.D.R., B.D., M.F., A.P.G., X.H., T.L., C.S.P., K.Ro., M.W.S. and M.E.Z. Data analysis: J.C.B., P.F.B., J.D.B., J.C., A.E.C., D.J.C., M.J.D., S.D.R., B.D., M.F., S.-C.F., A.P.G., X.H., L.K., J.K., Y.K., L.L., A.M., C.S.P., S.P., K.Ro., K.S., C.S., T.S., C.St., S.W., L.W. and M.E.Z. Contribution of samples, WES data or analytical tools: B.A., J.C.B., M.B., P.F.B., J.D.B., J.C., N.G.C., A.C., M.H.C., A.G.C., A.E.C., H.C., E.L.C., L.C., S.R.C., D.J.C., M.J.D., G.D., S.D.R., B.D., E.D., B.A.F., C.M.F., M.F., L.G., E.G., M.G., A.P.G., S.J.G., X.H., R.H., C.M.H., J.L.-L., P.J.G., H.K., S.M.K., L.K., A.K., J.K., Y.K., J.L., J.L., T.L., C.L., L.L., A.M., C.R.M., A.L.M., B.N., M.J.O., N.O., A.P., M.P., J.R.P., C.S.P., S.P., K.P., D.R., K.R., A.R., K.Ro., A.S., M.S., K.S., S.J.S., C.S., G.D.S., S.W.S., M.S.-R., T.S., P.S., D.S., M.W.S., C.St., J.S.S., P.S., K.T., O.V., A.V., S.W., C.A.W., L.W., L.A.W., J.A.W., T.W.Y., R.K.C.Y., M.E.Z. Writing of the paper: J.C.B., J.D.B., E.H.C., D.J.C., M.J.D., S.D.R., B.D., M.G., A.P.G., X.H., C.S.P., K.Ro., S.W.S., M.E.Z. Leads of ASC committees: J.D.B., E.H.C., M.J.D., B.D., M.G., K.Ro., M.W.S., J.S.S., M.E.Z. Administration of ASC: J.M.B.

Author Information New data included in this manuscript have been deposited at dbGAP merged with our published data under accession number phs000298.v1.p1 and is available for download at (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000298.v1.p1). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.B. (joseph.buxbaum@mssm.edu) or M.J.D. (mj Daly@broadinstitute.org).

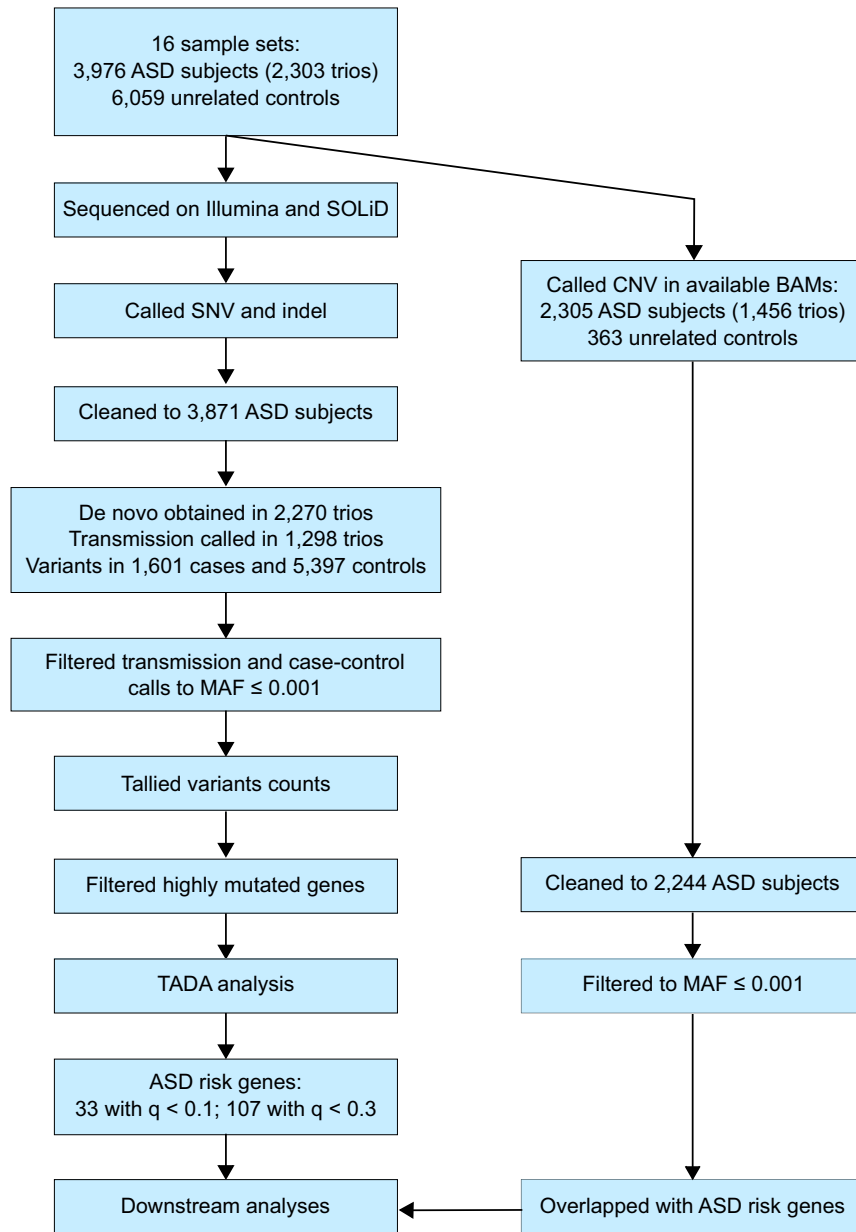
Silvia De Rubeis^{1,2}, Xin He³, Arthur P. Goldberg^{1,2,4}, Christopher S. Poultnery^{1,2}, Kaitlin Samocha⁵, A. Ericument Cicek³, Yan Kou^{1,2}, Li Liu⁶, Menachem Fromer^{2,4,5}, Susan Walker⁷, Tarjinder Singh⁸, Lambertus Klei⁹, Jack Kosmicki⁵, Shih-Chen Fu^{1,2}, Branko Aleksic¹⁰, Monica Biscaldi¹¹, Patrick F. Bolton¹², Jessica M. Brownfeld^{1,2}, Jinlu Cai^{1,2}, Nicholas G. Campbell^{13,14}, Angel Carracedo^{15,16}, Maria H. Chahrouh^{17,18}, Andreas G. Chiocchetti¹⁹, Hilary Coon^{20,21}, Emily L. Crawford^{13,14}, Lucy Crooks⁸, Sarah R. Currant¹², Geraldine Dawson²², Effthia Duketis¹⁹, Bridget A. Fernandez²³, Louise Gallagher²⁴, Evan Geller²⁵, Stephen J. Guter²⁶, R. Sean Hill^{17,18}, Iuliana Ionita-Laza²⁷, Patricia Jimenez Gonzalez²⁸, Helena Kilpinen²⁹, Sabine M. Klauk³⁰, Alexander Kolevzon^{1,2,31}, Irene Lee³², Jing Lei⁶, Terho Lehtimäki³³, Chiao-Feng Lin²⁵, Avi Ma’ayan³⁴, Christian R. Marshall⁷, Alison L. McInnes³⁵, Benjamin Neale³⁶, Michael J. Owen³⁷, Norio Ozaki¹⁰, Mara Parellada³⁸, Jeremy R. Parr³⁹, Shaun Purcell², Kaija Puura⁴⁰, Deepthi Rajagopalan⁷, Karola Rehnström⁸, Abraham Reichenberg^{1,2,41}, Aniko Sabo⁴², Michael Sachse¹⁹, Stephan J. Sanders⁴³, Chad Schafer⁶, Martin Schulte-Ruther⁴⁴, David Skuse^{32,45}, Christine Stevens³⁶, Peter Szatmari⁴⁶, Kristiina Tammimies⁷, Otto Valladares²⁵, Annette Voran⁴⁷, Li-San Wang²⁵, Lauren A. Weiss⁴³, A. Jeremy Willsey⁴³, Timothy W. Yu^{17,18}, Ryan K. C. Yuen³⁷, The DDD Study*, Homozygosity Mapping Collaborative for Autism*, UK10K Consortium*, The Autism Sequencing Consortium*, Edwin H. Cook²⁶, Christine M. Freitag¹⁹, Michael Gill²⁴, Christina M. Hultman⁴⁸, Thomas Lehner⁴⁹, Aarno Palotie^{5,50,51,52}, Gerard D. Schellenberg²⁵, Pamela Sklar^{2,4,53}, Matthew W. State⁴³, James S. Sutcliffe^{13,14}, Christopher A. Walsh^{17,18},

Stephen W. Scherer^{7,54}, Michael E. Zwick⁵⁵, Jeffrey C. Barrett⁸, David J. Cutler⁵⁵, Kathryn Roeder^{6,3}, Bernie Devlin⁹, Mark J. Daly^{1,7,36,56} & Joseph D. Buxbaum^{1,2,4,53,57,58}

¹Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ²Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York 10029, New York, USA. ³Ray and Stephanie Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. ⁴Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ⁵Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁶Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. ⁷Program in Genetics and Genome Biology, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada. ⁸The Wellcome Trust Sanger Institute, Cambridge, CB10 1SA, UK. ⁹Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA. ¹⁰Department of Psychiatry, Graduate School of Medicine, Nagoya University, Nagoya 466-8550, Japan. ¹¹Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, University Medical Center Freiburg; Center for Mental Disorders, 79106 Freiburg, Germany. ¹²Department of Child Psychiatry & SGDP Centre, King's College London Institute of Psychiatry, Psychology & Neuroscience, London, SE5 8AF, UK. ¹³Vanderbilt Brain Institute, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. ¹⁴Department of Molecular Physiology and Biophysics and Psychiatry, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA. ¹⁵Genomic Medicine Group, CIBERER, University of Santiago de Compostela and Galician Foundation of Genomic Medicine (SERGAS), 15706 Santiago de Compostela, Spain. ¹⁶Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah 21589, Kingdom of Saudi Arabia. ¹⁷Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁸Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts 02115, USA. ¹⁹Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt, 60528 Frankfurt, Germany. ²⁰Department of Internal Medicine, University of Utah, Salt Lake City, Utah 84132, USA. ²¹Department of Psychiatry, University of Utah, Salt Lake City, Utah 84108, USA. ²²Duke Institute for Brain Sciences, Duke University, Durham, North Carolina 27708, USA. ²³Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St John's, Newfoundland A1B 3V6, Canada. ²⁴Department of Psychiatry, School of Medicine, Trinity College Dublin, Dublin 8, Ireland. ²⁵University of Pennsylvania Perelman School of Medicine, Department of Pathology and Laboratory Medicine, Philadelphia, Pennsylvania 19104, USA. ²⁶Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois 60608, USA. ²⁷Department of Biostatistics, Columbia University, New York, New York 10032, USA. ²⁸Hospital Nacional de Niños Dr Saenz Herrera, CCSS, Child Developmental and Behavioral Unit, San José, Costa Rica. ²⁹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. ³⁰Division of Molecular Genome Analysis, German Cancer

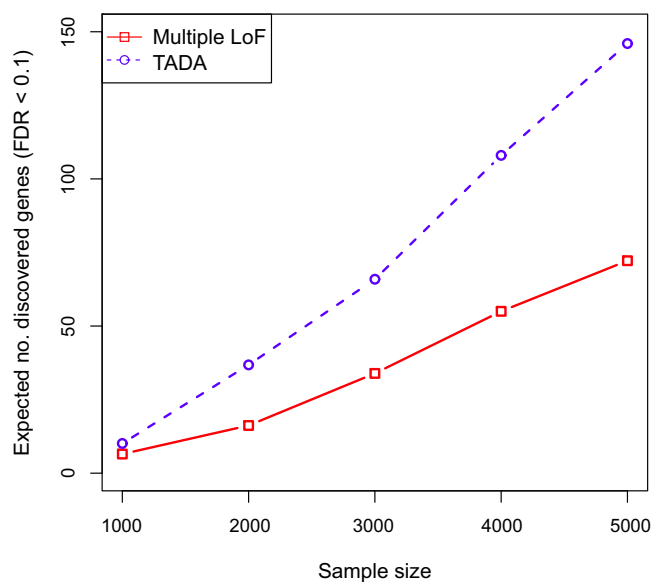
Research Center (DKFZ), 69120 Heidelberg, Germany. ³¹Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ³²Institute of Child Health, University College London, London, WC1N 1EH, UK. ³³Department of Clinical Chemistry, Fimlab Laboratories, SF-33100 Tampere, Finland. ³⁴Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ³⁵Department of Psychiatry Kaiser Permanente, San Francisco, California 94118, USA. ³⁶The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ³⁷MRC Centre for Neuropsychiatric Genetics and Genomics, and the Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff, CF24 4HQ, UK. ³⁸Child and Adolescent Psychiatry Department, Hospital General Universitario Gregorio Marañón, IISGM, CIBERSAM, Universidad Complutense, 28040 Madrid, Spain. ³⁹Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK. ⁴⁰Department of Child Psychiatry, University of Tampere and Tampere University Hospital, 33521 Tampere, Finland SF-33101. ⁴¹Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ⁴²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. ⁴³Department of Psychiatry, University of California at San Francisco, San Francisco, California 94143-0984, USA. ⁴⁴Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, Translational Brain Medicine in Psychiatry and Neurology, University Hospital RWTH Aachen / JARA Brain Translational Medicine, 52056 Aachen, Germany. ⁴⁵Department of Child and Adolescent Mental Health, Great Ormond Street Hospital for Children, National Health Service Foundation Trust, London, WC1N 3JH, UK. ⁴⁶Department of Psychiatry and Behavioural Neurosciences, Offord Centre for Child Studies, McMaster University, Hamilton, Ontario L8S 4K1, Canada. ⁴⁷Department of Child and Adolescent Psychiatry, Saarland University Hospital, D-66424 Homburg, Germany. ⁴⁸Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-171 77 Stockholm, Sweden. ⁴⁹National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892-9663, USA. ⁵⁰Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁵¹Institute for Molecular Medicine Finland, University of Helsinki, FI-00014 Helsinki, Finland. ⁵²Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁵³Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ⁵⁴McLaughlin Centre, University of Toronto, Toronto, Ontario M5S 1A1, Canada. ⁵⁵Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia 30322, USA. ⁵⁶Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁵⁷Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. ⁵⁸The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

*Lists of participants appear in the Supplementary Information.



Extended Data Figure 1 | Workflow of the study. The workflow began with 16 sample sets, as listed in Supplementary Table 1. DNA was obtained, and exomes were captured and sequenced. After variant calling, quality control was performed: duplicate subjects and incomplete families were removed and subjects with extreme genotyping, *de novo*, or variant rates were removed. Following cleaning, 3,871 subjects with ASD remained. Analysis proceeded

separately for SNVs and indels, and CNVs. *De novo* and transmission/non-transmission variants were obtained for trio data (published *de novo* variants from 825 trios^{11,13–15} were incorporated). This led to the TADA analysis, which found 33 ASD risk genes with an FDR < 0.1; and 107 with an FDR < 0.3. CNVs were called in 2,305 ASD subjects. BAM, binary alignment/map; MAF, minor allele frequency.



Extended Data Figure 2 | Expected number of ASD genes discovered as a function of sample size. The multiple LoF test (red) is a restricted version of TADA that uses only the *de novo* LoF data. TADA (blue) models *de novo* LoF, *de novo* Mis3, LoF variants transmitted/not transmitted and LoF variants observed in case-control samples. The sample size (n) indicates either n trios for which we record *de novo* and transmitted variation (TADA), or n trios for which we record only *de novo* events (multiple LoF), plus n cases and n controls.



Extended Data Figure 3 | Heat map of the numbers of variants used in TADA analysis from each data set in genes with an FDR < 0.3. Left, variants in affected subjects; right, unaffected subjects. For the counts, we only included *de novo* LoF and Mis3 variants, transmitted/untransmitted and case-control

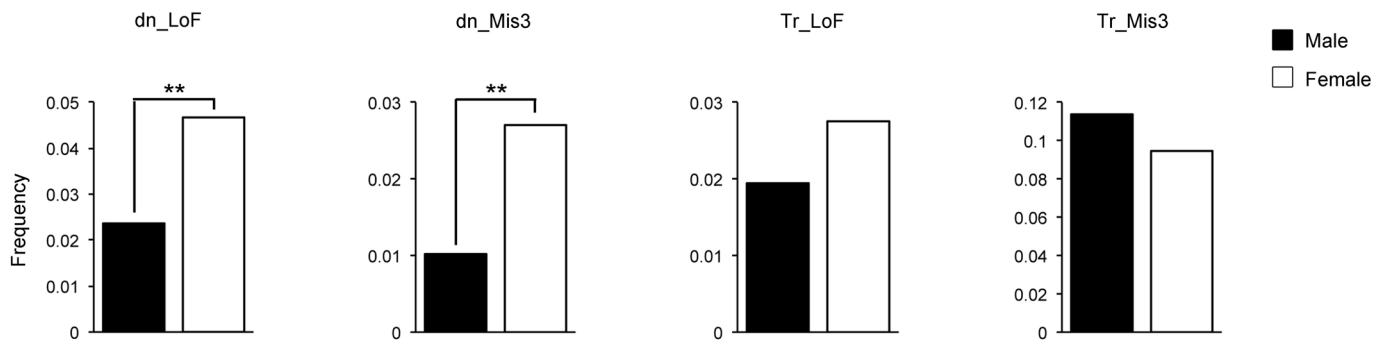
LoF variants. These variant counts are normalized by the length of coding regions of each gene and sample size of each data set ($(|trio| + |case|)$ for the left, $(|trio| + |control|)$ for the right). Description of the samples can be found in Supplementary Table 1.



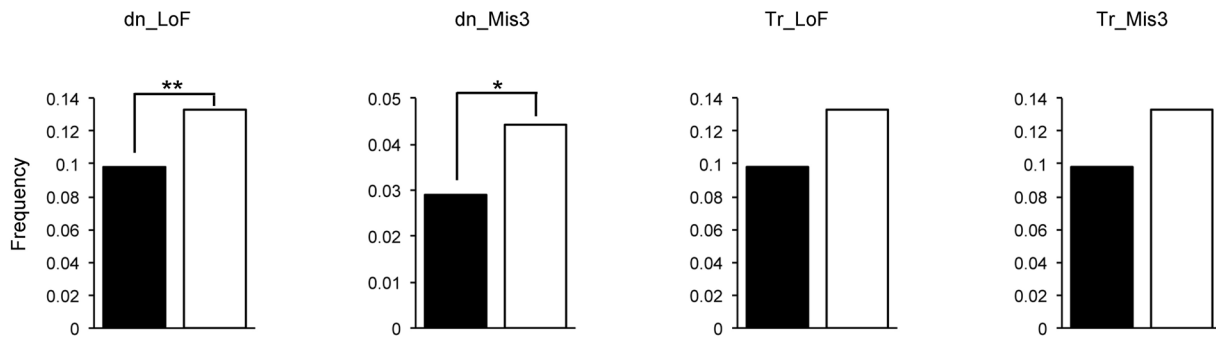
Extended Data Figure 4 | Genome browser view of the CNV deletions identified in ASD-affected subjects. The deletions are displayed in red if with unknown inheritance, in grey if inherited, and in black in unaffected subjects.

Deletions in parents are not shown. For deletions within a single gene, all splicing isoforms are shown.

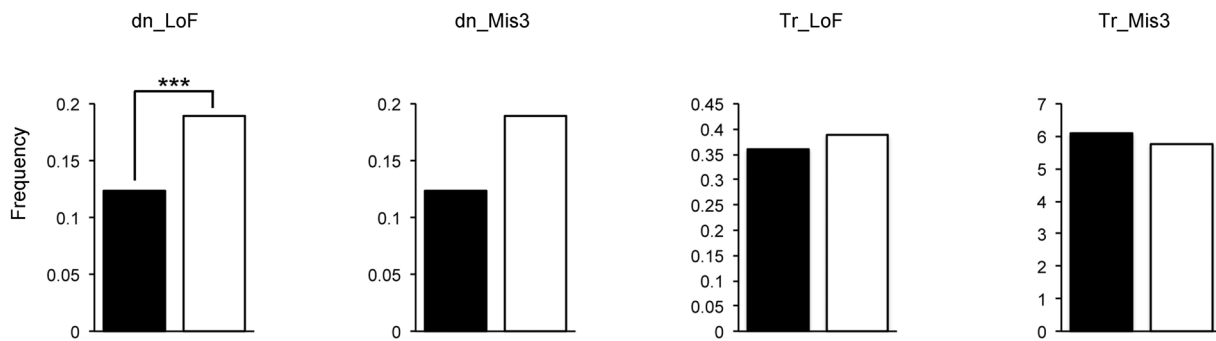
Genes $q < 0.1$



Genes $q < 0.3$

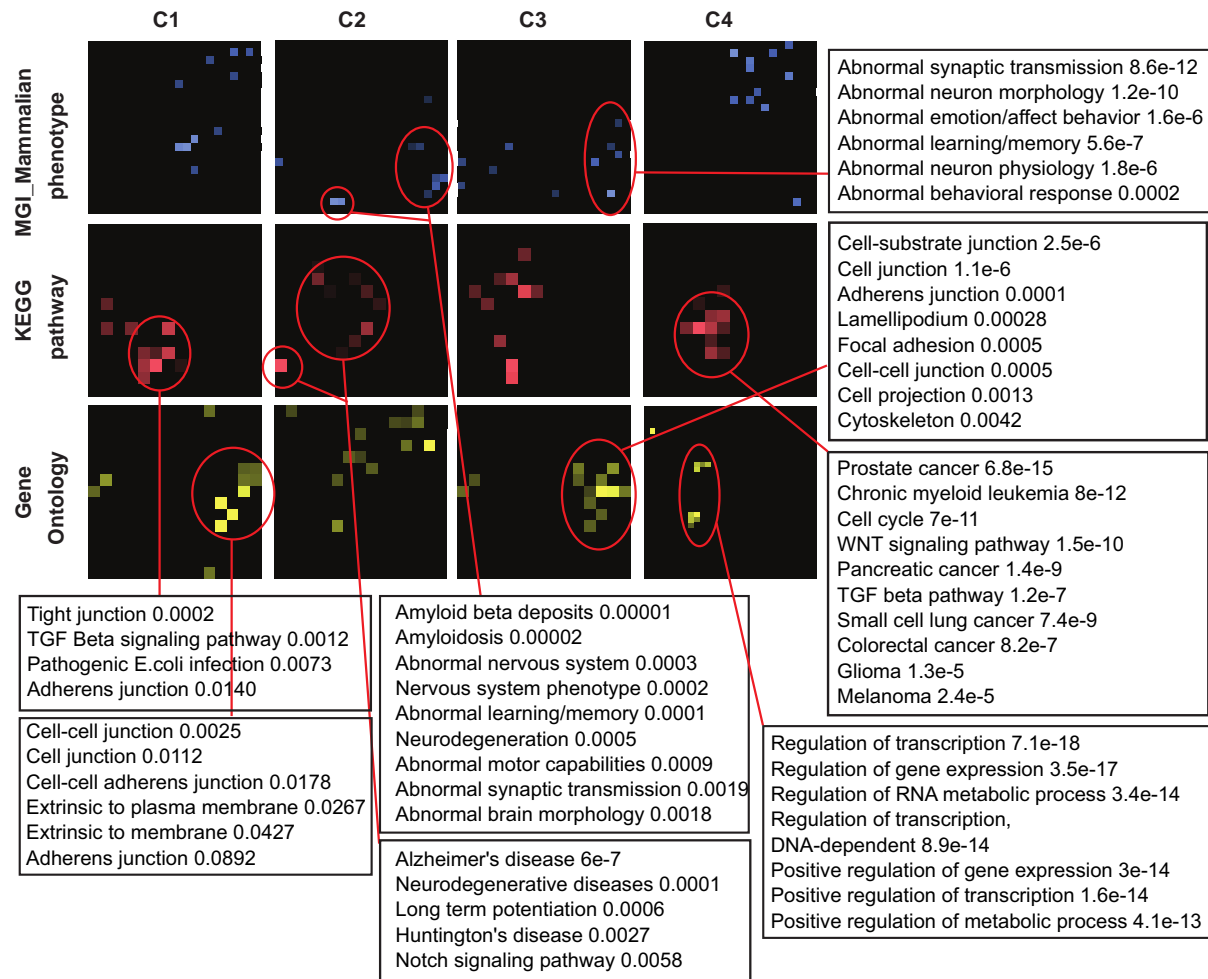


All TADA genes



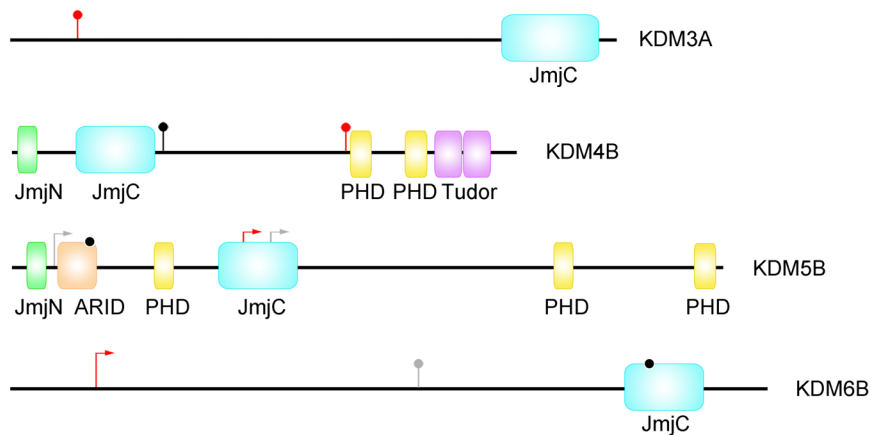
Extended Data Figure 5 | Frequency of variants by gender. Frequency of *de novo* (dn) and transmitted (Tr) variants per sample in males (black) and females (white) for genes with an FDR < 0.1 (top row), FDR < 0.3 (middle

row), or all TADA genes (bottom row). The P values were determined by one-tailed permutation tests (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

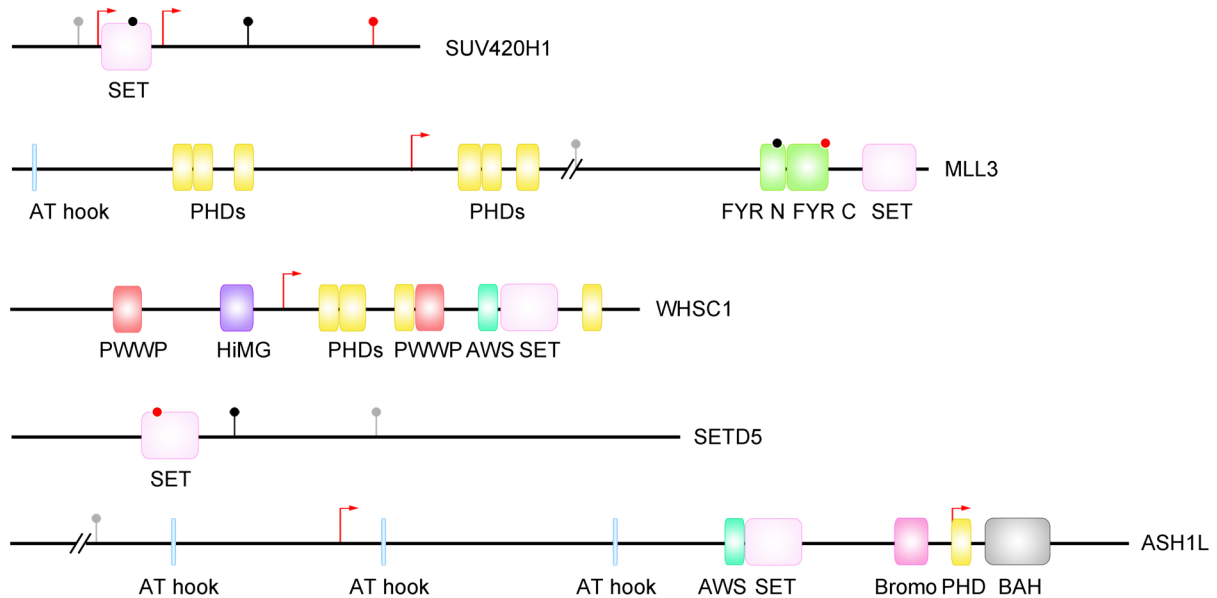


Extended Data Figure 6 | Enrichment terms for the four clusters identified by protein-protein interaction networks. *P* values calculated using mouse-genome-informatics-mammalian-phenotype (MGI_Mammalian phenotype, blue), Kyoto encyclopaedia of genes and genomes (KEGG) pathways (red), and gene ontology biological processes (yellow) are indicated.

Histone demethylases

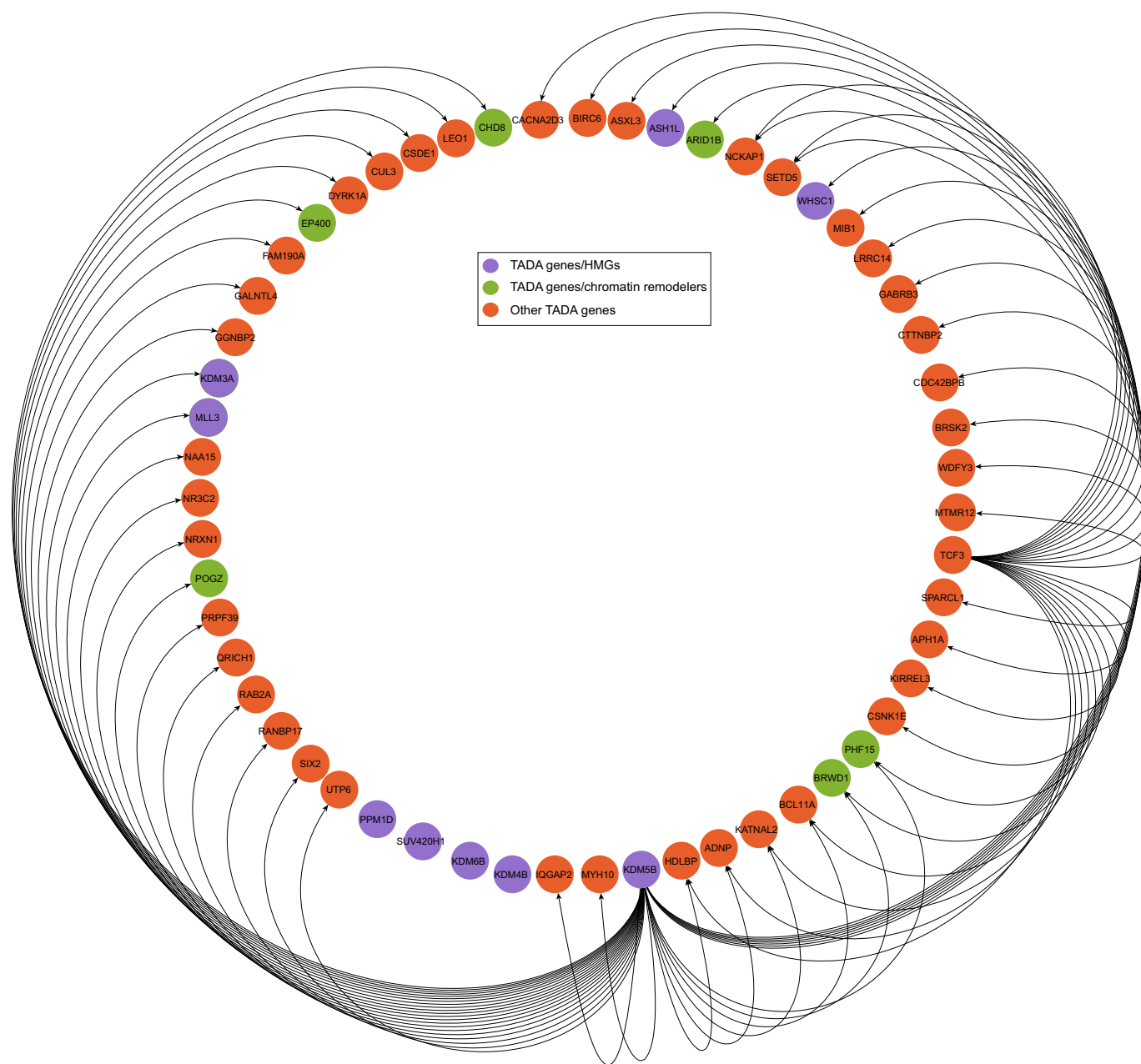


Histone methyltransferases



Extended Data Figure 7 | *De novo* variants in SET lysine methyltransferases and jumonji lysine demethylases. Mis3 variants are in black, LoF in red, and variants identified in other disorders in grey (Fig. 5). ARID, AT-rich interacting domain; AWS, associated with SET domain; BAH, bromo adjacent homology;

bromo, bromodomain; FYR C, FY-rich C-terminal domain; FYR N, FY-rich N-terminal domain; HiMG, high mobility group box; JmjC, jumonji C domain; JmjN, jumonji N domain; PHD, plant homeodomain; PWWP, Pro-Trp-Trp-Pro domain; SET, Su(var)3-9, enhancer-of-zeste, trithorax domain.



Extended Data Figure 8 | Transcription regulation network of TADA genes only. Edges indicate transcription regulators (source nodes) and their gene targets (target nodes) based on the ChEA network.

Extended Data Table 1 | CNVs hitting TADA genes

Gene	ASD subject		Unaffected parent*			Unaffected	Odds Ratio†
	Unknown Inheritance	Inherited	Tr-ASD	NT	Tr-not-ASD		
q-value < 0.1							
ANK2	1						∞
ASXL3	1						∞
VIL1		1	1				1.49
0.1 ≤ q-value < 0.3: Evidence for role in ASD							
UTP6	1						∞
DNAH10		1	1				1.49
ATP1B1	1						∞
GGNBP2	1						∞
NRXN1		2	1				2.99
WHSC1	1						∞
HDLBP‡	1	2	1		1	1	2.24
CERS4		1	1				1.49
SHANK3	4						∞
IQGAP2	1						∞
0.1 ≤ q-value < 0.3: Evidence against role in ASD							
EP400						1	0
SLCO1B1 ‡ §	1	1	1	1		1	0.996
SLCO1B3 §		1	1	2		1	0.37
KDM6B						1	0

Count of deletion CNVs inferred from sequence for ASD subjects and those unaffected by ASD. Number of subjects and family status: 849 ASD subjects without family information; 1,467 ASD subjects in families; 2,766 unaffected parents; 319 unaffected siblings of ASD subjects; 373 unaffected subjects without family information. NT, parent a carrier but CNV not transmitted to affected child; Tr-ASD, transmitted to ASD subject from carrier parent; Tr-not-ASD, parent transmits a CNV to an unaffected child.

*No parents in this count were affected; seven parents in the study were affected, none carried a CNV reported in the table and these subjects did not enter the calculation.

†To compute the odds ratio we count the number of affected carriers (a), unaffected carriers (including parents) (b), affected subjects who do not have the CNV (c), and unaffected non-carriers (d). The odds ratio = (ad)/(bc).

‡One parent transmits the CNV to an affected and unaffected offspring; to obtain the total count of controls with a CNV, subtract one.

§ Genes are adjacent in the genome (see Extended Data Fig. 4). For three subjects both genes are affected by the same CNV (1 ASD and 2 unaffected subjects).

The contribution of *de novo* coding mutations to autism spectrum disorder

Ivan Iossifov^{1*}, Brian J. O’Roak^{2,3*}, Stephan J. Sanders^{4,5*}, Michael Ronemus^{1*}, Niklas Krumm², Dan Levy¹, Holly A. Stessman², Kali T. Witherspoon², Laura Vives², Karynne E. Patterson², Joshua D. Smith², Bryan Paepers², Deborah A. Nickerson², Jeanselle Dea⁴, Shan Dong^{5,6}, Luis E. Gonzalez⁷, Jeffrey D. Mandell⁴, Shrikant M. Mane⁸, Michael T. Murtha⁷, Catherine A. Sullivan⁷, Michael F. Walker⁴, Zainulabedin Waqar⁷, Liping Wei^{6,9}, A. Jeremy Willsey^{4,5}, Boris Yamrom¹, Yoon-ha Lee¹, Ewa Grabowska^{1,10}, Ertugrul Dalkic^{1,11}, Zihua Wang¹, Steven Marks¹, Peter Andrews¹, Anthony Leotta¹, Jude Kendall¹, Inessa Hakker¹, Julie Rosenbaum¹, Beicong Ma¹, Linda Rodgers¹, Jennifer Troge¹, Giuseppe Narzisi^{1,10}, Seungtae Yoon¹, Michael C. Schatz¹, Kenny Ye¹², W. Richard McCombie¹, Jay Shendure², Evan E. Eichler^{2,13}, Matthew W. State^{4,5,7,14} & Michael Wigler¹

Whole exome sequencing has proven to be a powerful tool for understanding the genetic architecture of human disease. Here we apply it to more than 2,500 simplex families, each having a child with an autistic spectrum disorder. By comparing affected to unaffected siblings, we show that 13% of *de novo* missense mutations and 43% of *de novo* likely gene-disrupting (LGD) mutations contribute to 12% and 9% of diagnoses, respectively. Including copy number variants, coding *de novo* mutations contribute to about 30% of all simplex and 45% of female diagnoses. Almost all LGD mutations occur opposite wild-type alleles. LGD targets in affected females significantly overlap the targets in males of lower intelligence quotient (IQ), but neither overlaps significantly with targets in males of higher IQ. We estimate that LGD mutation in about 400 genes can contribute to the joint class of affected females and males of lower IQ, with an overlapping and similar number of genes vulnerable to contributory missense mutation. LGD targets in the joint class overlap with published targets for intellectual disability and schizophrenia, and are enriched for chromatin modifiers, FMRP-associated genes and embryonically expressed genes. Most of the significance for the latter comes from affected females.

Autism spectrum disorder (ASD) is characterized by impaired social interaction and communication, repetitive behaviour and restricted interests. It has a strong male bias, especially in high-functioning affected individuals. The contribution from transmission has long been suspected from increased sibling risk¹, but more recently the role of germline *de novo* (DN) mutation has been established, first from large-scale copy number variation in simplex families^{2–5}, and subsequently from exome sequencing. The smaller DN variants observed by DNA sequencing pinpoint candidate gene targets^{6–8}. These developments have promoted a new model for causation, and re-evaluation of sibling risk^{9,10}.

Here we report whole exome sequencing of the Simons Simplex Collection (SSC)¹¹ and an extensive list of DN mutated targets, including 27 recurrent LGD (nonsense, frameshift and splice site) targets. The size and uniformity of this study allow an unprecedented evaluation of genetic vulnerability to ASD. We subdivide target sets by mutation type (missense and LGD) and affected child status (gender and non-verbal IQ, to which we refer throughout as ‘IQ’), and explore the overlap between target sets and their enrichment for certain gene categories. We make estimates of the number of genes vulnerable to a given mutation type and the proportion of simplex autism resulting from DN mutation for each affected subpopulation.

SSC sequencing and validation

We report on 2,517 of ~2,800 SSC families including ~800 that were previously published^{6–8}. We sequenced 2,508 affected children, 1,911

unaffected siblings and the parents of each family. Within the SSC, the overall gender bias in affected individuals, 7 males to 1 female, is nearly twice that typically reported. Exomes were analysed at Cold Spring Harbor Laboratory (CSHL), Yale School of Medicine, and University of Washington (Extended Data Figs 1 and 2 and Supplementary Table 1). Pipelines were blind with respect to affected status. For uniformity, all data were reanalysed with the CSHL pipeline, allowing comparison of analysis tools. All calls were validated or strongly supported, as listed (Supplementary Table 2) and described (Methods).

Rates and targets of DN mutation

For greatest precision we measured DN rates in quad families (one affected and one unaffected child) over genomic positions at which all family members had $\geq 40\times$ sequence coverage (Methods and Supplementary Table 3). This ‘joint $40\times$ region’ in the SSC was 32 gigabases (Gb) in total, or 48% of the targeted exome, from 1,867 quads. DN events were shared by siblings 1% of the time (Supplementary Table 2); and 1% of mutations had nearby nucleotide positions altered, presumably by single mutagenic events^{12–14} (Supplementary Table 4). The overall rate of base substitution is 1.8×10^{-8} ($\pm 10^{-9}$) per base pair (Supplementary Table 5).

Rates of DN synonymous mutation in affected (0.34 per child) and unaffected (0.33 per child) siblings do not differ significantly (Fig. 1). By contrast, LGD mutations occur at significantly higher rates in affected versus unaffected siblings (Fig. 1 and Extended Data Fig. 3). The rate of LGD mutations is 0.12 in unaffected siblings and 0.21 in affected

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. ²Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA. ³Molecular & Medical Genetics, Oregon Health & Science University, Portland, Oregon 97208, USA. ⁴Department of Psychiatry, University of California, San Francisco, San Francisco, California 94158, USA. ⁵Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520, USA. ⁶Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China. ⁷Child Study Center, Yale University School of Medicine, New Haven, Connecticut 06520, USA. ⁸Yale Center for Genomic Analysis, Yale University School of Medicine, New Haven, Connecticut 06520, USA. ⁹National Institute of Biological Sciences, Beijing 102206, China. ¹⁰New York Genome Center, New York, New York 10013, USA. ¹¹Department of Medical Biology, Bulent Ecevit University School of Medicine, 67600 Zonguldak, Turkey. ¹²Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York 10461, USA. ¹³Howard Hughes Medical Institute, Seattle, Washington 98195, USA. ¹⁴Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06520, USA.

*These authors contributed equally to this work.

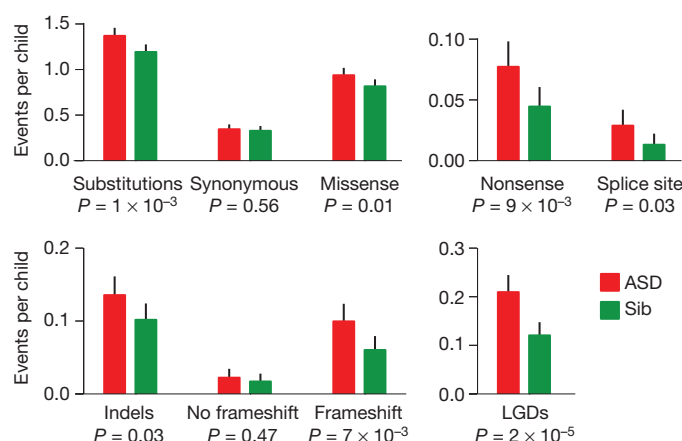


Figure 1 | Rates of *de novo* events by mutational type in the SSC. Rates per child are estimated from the 40× joint coverage target region, then extrapolated for the entire exome. Mutation types are displayed by class, and the combined rate for all LGDs is shown at the bottom right. For each event type, the significance between probands and unaffected siblings is given. Sib, unaffected siblings. The errors bar represent 95% confidence interval for the mean rates.

probands, an ‘ascertainment differential’ of $0.21 - 0.12 = 0.09$ ($P = 2 \times 10^{-5}$). Thus, we estimate ~43% (0.09 out of 0.21) of LGD events in probands contribute to ASD diagnoses. For DN missense, the rate is 0.82 for unaffected siblings and 0.94 for affected probands, an ascertainment differential of 0.12 ($P = 0.01$). We estimate only ~13% (0.12 out of 0.94) of DN missense events in probands contribute to ASD diagnoses. There is a wide confidence interval for the missense ascertainment differential (Supplementary Table 6); for this reason, we consider primarily the LGD events for our analysis and look on missense data as supporting.

To identify gene targets for DN mutation, we examined all family data including trios. We provide a complete list of all mutations (Supplementary Table 2) along with the number of mutations of each type in each gene (Supplementary Table 7). A total of 391 DN LGD mutations in 353 target genes were identified and validated in autism probands. Of these, 27 target genes were recurrent (Fig. 2). Among 1,500 missense targets in probands, 145 were recurrent.

We examined all alleles transmitted opposite a DN LGD target. We saw no instance in 391 observations in which the allele opposite an LGD target carried a rare transmitted LGD variant (in <1% of parental exomes), and only four in which such an allele carried a rare missense variant. Thus, the DN mutations do not generally cause homozygous loss-of-function of their target (Supplementary Table 8).

Confirming previous results^{7,8,15}, observed DN mutations arise three times as often in the paternal background, and mutation rates rise with age of either parent (Extended Data Fig. 4 and Methods). The latter may provide a partial explanation for increased autism rates in children born of older parents.

Functional clustering in target genes

Previous studies presented evidence of functional clustering in targets of DN LGD mutation in affected individuals^{6–8,16}. Our larger data set was examined with an improved null ‘length model’ for mutation in which the probability of DN mutation in a gene is proportional to its length (Methods and Extended Data Fig. 5). We tested for enrichment within DN LGD and missense targets in probands and siblings for the following six classes: (1) FMRP target genes, with transcripts bound by the fragile X mental retardation protein^{8,17}; (2) genes encoding chromatin modifiers; (3) genes expressed preferentially in embryos^{18,19}; (4) genes encoding postsynaptic density proteins²⁰; (5) essential genes²¹; and (6) genes identified as Mendelian disease genes²² (Table 1, Supplementary Table 6 and Methods). These data provide the strongest evidence yet for

overlap of DN LGD targets in affected probands with FMRP targets (55 observed versus 34.1 expected; $P = 4 \times 10^{-4}$) and chromatin modifiers (26 observed versus 11.8 expected; $P = 3 \times 10^{-4}$). We also observed signal from mutation in genes expressed in embryonic development²³ (65 observed versus 45.0 expected; $P = 2 \times 10^{-3}$). The latter signal comes mainly from the small number of female affected individuals (23 observed versus 8.5 expected from 67 LGD targets; $P = 5 \times 10^{-6}$). The 27 genes with recurrent LGDs show strong enrichment for FMRP targets (14 observed versus 2.6 expected; $P = 4 \times 10^{-8}$) and chromatin modifiers (6 observed versus 0.9 expected; $P = 2 \times 10^{-4}$). By contrast, no significant enrichment for these gene sets is seen for the DN LGD targets in unaffected siblings.

The 1,500 DN missense targets in probands are also enriched for FMRP targets and embryonically expressed genes. We observe 171 FMRP targets (144.8 expected; $P = 0.03$), and 220 embryonically expressed genes (191.4 expected; $P = 0.03$). As before, the signal for embryonically expressed genes comes almost entirely from the small number of female affected individuals (48 observed, 31.1 expected from 244 targets; $P = 0.002$). With the exception of chromatin modifiers, contributory DN missense and LGD mutations tend to strike similar functional classes of genes.

De novo mutation and IQ

Higher IQ probands are heavily skewed towards males²⁴. For further analyses, we chose to divide the affected male population roughly in half into higher and lower IQ sets. We investigated whether higher IQ (>90) males comprise a population with a distinguishable genetic signature. There is a decreased ascertainment differential for DN LGD mutations in male children with higher IQ relative to other affected individuals (Extended Data Fig. 3 and Supplementary Table 6). This is not statistically significant over the joint 40× region. However, over the entire data set, the drop in IQ is 5 points for males with DN LGD mutation compared to those without mutation ($P = 0.01$; Fig. 2). The mean IQ of affected males with recurrent DN LGDs drops 20 points ($P = 0.00001$, Fig. 2). Significance is also evident as we examine targets by functional class. Males with LGD mutations in FMRP targets have an average 14-point drop ($P = 0.001$). This trend continues with LGD targets in the other functional classes—chromatin modifiers and embryonically expressed genes—but with reduced significance. We observe little signal from DN missense mutation, even in recurrent targets, either because these events are less likely to contribute or because they are less severe. Female probands show the same trends as males, but as they comprise a smaller population, the significance is weak (Fig. 2).

Further evidence for a distinguishable signature among the higher IQ comes from the functional enrichment within DN target gene sets. LGD targets in females are enriched for all three functional gene classes. LGD targets in lower IQ affected males are significantly enriched for the FMRP-associated and chromatin-modifier gene classes (Supplementary Table 6). However, for LGD targets in higher IQ males we see no statistically significant enrichment for any of the gene categories.

Target overlaps in children and mutation-type groups

We partitioned children into four primary groups: unaffected siblings, affected females, affected males with higher IQ, and affected males with lower IQ. We analysed these and various combinations for three types of DN mutations: LGDs, missense and synonymous (Supplementary Table 6). Targets of synonymous mutations in all children and targets of LGD and missense mutations in unaffected siblings have no significant overlap with targets from any other group. We see no significant overlap between targets in higher IQ males with targets from other groups. In strong contrast, the 67 LGD targets from affected females overlap significantly with the 166 LGD targets from lower IQ affected males (10 observed, 1.3 expected, $P = 7 \times 10^{-7}$). We therefore refer to the group of lower IQ males and affected females as a ‘joint’ class. In this class, the 874 missense and 223 LGD targets also overlap significantly (39 observed, 22.1 expected, $P = 0.0008$). Thus, not only do missense and

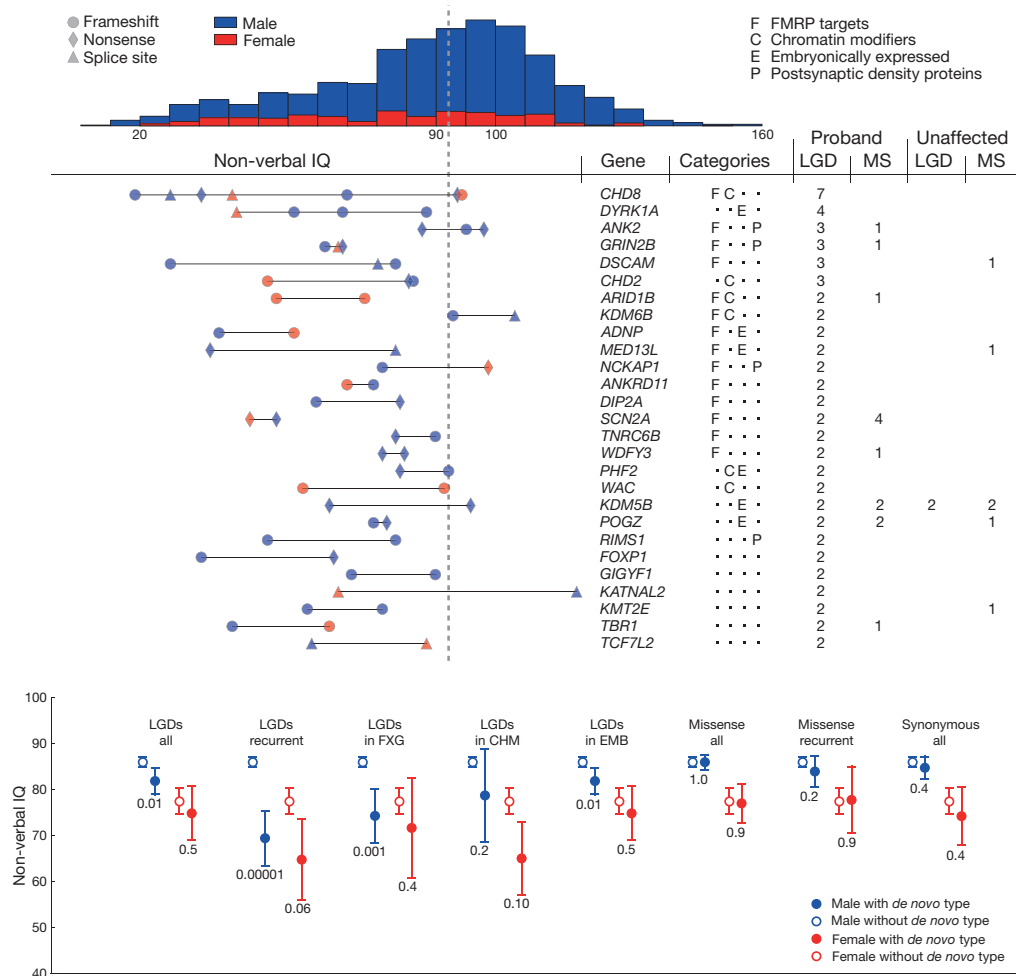


Figure 2 | Recurrently hit genes and non-verbal IQ. Affected females account for 13.5% of the SSC with a mean IQ of 78, whereas affected males have a mean IQ of 86 (top, $P = 10^{-7}$ by Student's t -test). Vertical dashed line indicates an IQ of 90. Middle (left) shows IQ for affected children with LGD mutations in genes hit recurrently (right). Recurrently mutated genes are clustered into four categories as shown. The last four columns give overall numbers of DN LGD and missense (MS) mutations. Bottom, eight classes of

DN mutations are considered: all LGDs, recurrent LGDs, LGDs in FMRP targets (FXG), LGDs in chromatin modifiers (CHM), LGDs in embryonically expressed genes (EMB), all missense mutations, recurrent missense mutations and synonymous mutations. Probands are divided by the presence of DN mutations and gender. Means, 95% confidence intervals and P values (Student's t -test) are shown.

LGD mutation target genes with shared functionality, the same genes are sometimes targeted.

Number of vulnerable genes

Our analysis of functional clustering and overlaps within target classes suggests that the mutations ascertained in probands target restricted

sets of vulnerable genes. We next sought evidence for excess recurrence of targets. We first examined synonymous mutations and mutations in unaffected children. Among the 647 synonymous events in probands, there are 25 gene targets found in more than one child, close to the null expectation of 19.9 ($P = 0.13$). Recurrent LGD ($n = 3$ out of 179 events) or missense targets (70 out of 1,143 events) in unaffected siblings are

Table 1 | Enrichment of DN mutations in six gene classes

Gene class	No. of genes	rDN LGD (ASD)			DN LGD (ASD)			DN miss (ASD)			DN LGD (sib)			DN miss (sib)		
		Overlap (27)			Overlap (353)			Overlap (1,513)			Overlap (176)			Overlap (1,066)		
		Obs	Exp	P	Obs	Exp	P	Obs	Exp	P	Obs	Exp	P	Obs	Exp	P
FMRP	842	14	2.6	4×10^{-8}	55	34.1	4×10^{-4}	171	144.8	0.03	14	17.0	0.52	117	102.9	0.15
Chromatin	428	6	0.9	2×10^{-4}	26	11.8	3×10^{-4}	57	50.0	0.31	5	5.9	1.00	37	35.6	0.80
Embryonic	1,912	6	3.4	0.15	65	45.0	2×10^{-3}	220	191.4	0.03	20	22.5	0.65	142	136.0	0.58
PSD	1,445	4	2.5	0.31	34	32.5	0.78	159	138.1	0.07	22	16.2	0.15	113	98.1	0.12
Essential	1,750	7	3.2	0.04	50	42.4	0.22	201	180.3	0.10	20	21.2	0.91	127	128.1	0.96
Mendelian	256	0	0.6	1.00	3	8.0	0.07	31	34.0	0.66	5	4.0	0.61	20	24.1	0.47
DN LGD (Scz)	93	2	0.3	0.03	9	3.7	0.01	16	15.7	0.90	2	1.8	0.71	8	11.2	0.45
DN LGD (ID)	30	3	0.1	1×10^{-4}	8	1.2	3×10^{-5}	10	4.9	0.04	0	0.6	1.00	5	3.5	0.41

We tested eight classes (Methods) for enrichment against five lists of targets of DN mutations. These include genes with (1) recurrent DN LGD mutations in probands (rDN LGD (ASD)); (2) DN LGD mutations in probands (DN LGD (ASD)); (3) DN missense mutations in probands (DN miss (ASD)); (4) DN LGDs in siblings (DN LGD (sib)); and (5) DN missense mutations in siblings (DN miss (sib)). Observed (obs) and expected (exp) numbers are shown with P values obtained from two-sided binomial tests. Expected numbers and P values are based on a length model in which DN mutations occur randomly in all genes, proportional to length.

also close to null expectations ($P = 0.2$ and 0.04 , respectively). In affected males with higher IQ there are no excess recurrent targets among 137 LGDs mutations (2 observed, 1.0 expected, $P = 0.3$) or among 728 missense mutations (26 observed, 24.7 expected, $P = 0.4$). By contrast, among probands the number of recurrent LGD ($n = 27$ out of 391 events) and missense targets (145 out of 1,675 events) are not compatible with the null expectation of 7.6 ($P < 0.0001$) and 115.0 ($P = 0.001$), respectively. Given these findings, as well as the lack of overlap between targets of higher and lower IQ males, we focused on the joint class of female probands and affected males of lower IQ. For the joint class, there were 22 recurrent LGD targets among 254 events with 3.3 expected ($P < 0.0001$). For the 944 missense events, 60 recurrent targets are observed with 40.2 expected ($P = 0.0005$).

We next used recurrence analysis and the length model to estimate the number of vulnerable genes (Fig. 3) and the probability that a recurrent mutation of a given type is contributory (Methods). The most likely number of genes vulnerable to DN mutations in the joint class is estimated to be 387 for LGD targets with a 95% confidence interval of 149–915, and 404 for DN missense targets (confidence interval 71–3,050). From the length model and our estimate that 43% of LGD mutations are contributory, we have 90% confidence that a given LGD mutation contributes to autism in a gene recurrently hit by an LGD mutation (Methods). By the same methods, we compute 35% confidence in contribution from missense mutations in recurrent targets. Using existing models for prioritizing targets⁷, we list all targets of recurrent DN coding mutation according to their rank (Supplementary Table 9).

Discussion

The SSC was assembled with the explicit hypothesis that finding targets of DN mutation would be a path to gene discovery. We now have 353 candidate LGD gene targets, 27 genes recurrently hit by LGD events, and 145 recurrent missense targets, each with about 40%, 90% and 35% chance of being contributory, respectively.

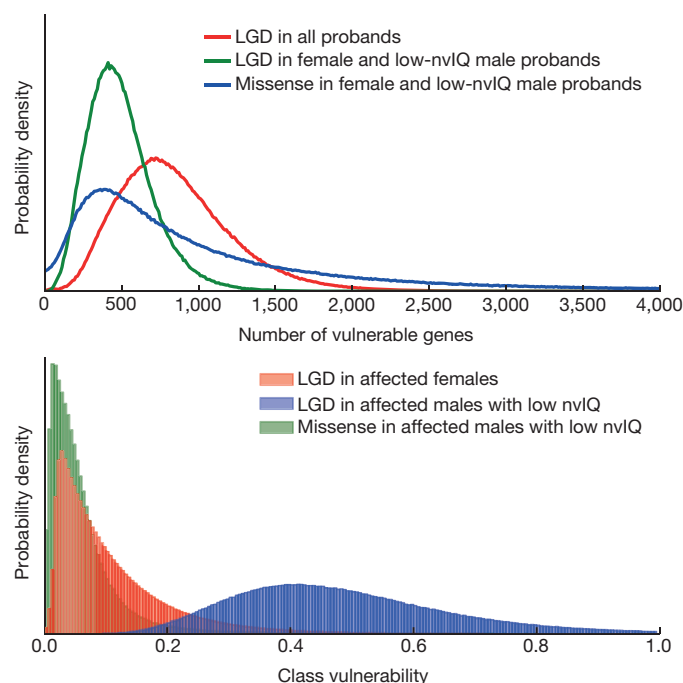


Figure 3 | Number of vulnerable genes and class vulnerability. We assume the property of being vulnerable is independent of gene length, but the probability of being hit by mutation is proportional to gene length. We use the observed rates of mutation of a given type in specified populations and number of recurrent mutations to estimate the number of genes vulnerable to those mutations (top). The degrees of vulnerability in those classes are the distributions shown in the bottom panel (Methods). nvIQ, non-verbal IQ.

We use the ascertainment differential as an estimate of contribution. The sum of the ascertainment differentials for missense, nonsense, consensus splice site disruption and frameshift DN mutations is 0.21 per affected child. Adding 0.06, the ascertainment differential from large DN copy number variants^{2,3}, brings the total to 0.27 (Fig. 4). Excluding higher IQ males, the value is 0.33. In affected females it is 0.45. This is a conservative estimate for the role of DN mutation in the SSC families because we have not yet ascertained intermediate-size DN copy number variants (CNVs), copy-neutral rearrangements, regulatory mutations or mutations of non-coding genes.

Although the SSC is a simplex collection, it is probably only marginally depleted for high-risk families because small brood size prevents the birth of several affected children, especially if the unaffected sibling is female. We estimate¹⁰ and confirm⁹ by gender bias in unaffected siblings (1,400 females and 1,264 males, $P = 0.0089$) that ~40% of the SSC families are high-risk. In a simple genetic model, DN mutation has no role in high-risk families but is obligatory for low-risk families¹⁰, so DN mutation would contribute to ~60% of the SSC. The sum of the ascertainment differential for all observable DN types in all the probands is about 30%, about half of that. If the number of unobserved and consequential DN mutations is similar to the number of observed and consequential DN exome mutations, the actual contribution is not far from that predicted by this simple model.

Targets and cognitive defects

We examined the incidence and targets of DN LGD mutations for children with lower and higher IQs. Affected children with higher IQs have a greater incidence of LGD mutations than unaffected siblings, but a lower incidence than affected females or males with lower IQ. Moreover, there are few recurrently hit genes among the DN LGD targets of affected males with higher IQ, and little overlap with the DN LGD targets of affected males with lower IQ or females. LGD targets in higher IQ males are not enriched for the FMRP-associated genes. These

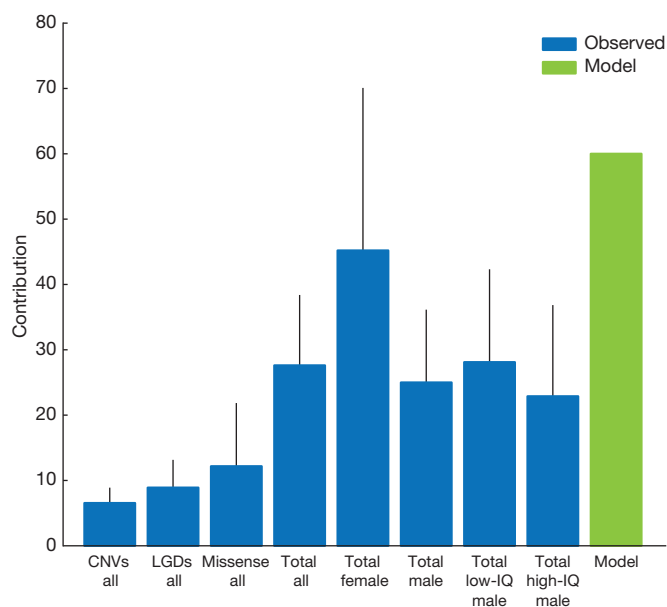


Figure 4 | Estimated contributions of CNVs, LGDs and missense DN mutations to simplex ASD. Ascertainment differentials for three types of DN mutation (CNVs, LGDs and missense) are interpreted as a measure of 'contribution,' the percentage of probands in whom the mutation contributed to diagnosis. We combine the three mutation types in 'total' on the assumption of additivity. We present this measure for 'all' probands and selected subpopulations as indicated. We also show the expected contribution of all DN mutation in a simplex collection computed from a simple genetic model⁹ ('model'). Error bars represent 95% confidence intervals.

observations suggest a different distribution of genetic mechanisms underlying ASD in higher IQ males.

We can examine overlap between LGD targets for autism, with published targets for intellectual disability and schizophrenia^{25–29}. We applied our length model for mutation incidence and found significant overlap of intellectual disability and schizophrenia targets with ASD targets (Table 1), but only in the joint class of affected males with lower IQ and females (Supplementary Table 6). The overlap can have many explanations: diagnostic conflation; pleiotropy for the same mutation; different consequences for different mutations in the same gene; and varying genetic or environmental background. The DN targets of affected males with higher IQ do not overlap these sets, again suggesting distinct mechanisms.

Properties of target classes

This study is sufficiently large and uniform to enable inferences about targets, distinguished by mutation types, properties of affected children and target functions. We observe a significant contribution from missense mutations, with an overall magnitude comparable to that from LGD mutations. Both LGD and missense mutation targets are enriched in the same functional gene sets, especially among lower IQ males (Supplementary Table 6). Excluding higher IQ males, we estimate the most likely number of genes vulnerable to LGDs is about 400, with a similar number of genes vulnerable to missense. The two sets overlap substantially.

Targets in autism are enriched in certain functional categories, providing deeper support for previously published observations^{6–8}. FMRP-associated genes and chromatin modifiers are prominent targets in all groups except higher IQ males. The former are thought to function in neuroplasticity. Embryonically expressed genes are significantly enriched as LGD or missense targets, but only in females. Enrichment in these genes may reflect that these contributory mutations cause alterations before a female protective effect takes place.

Recurrent LGD targets encode receptors, ion channels and synaptic proteins likely to function directly in neuro-circuitry (for example, *SCN2A*, *GRIN2B* and *RIMS1*), but also proteins functioning in cytoskeletal remodelling (for example, *ANK2* and *MED13L*) and transcriptional regulation. Chromodomain helicase gene family members carry many recurrent LGDs. The most frequently hit gene is *CHD8* (ref. 30), followed by *CHD2* (three LGDs) and four other members (one LGD each) of that family. *CHD8* is a transcriptional regulator thought to be important for suppression of the Wnt- β -catenin signalling pathway through histone H1 recruitment³¹. Another intriguing target is the protein kinase *DYRK1A*, hit four times and located in the Down's syndrome critical region⁷.

Gene vulnerability and molecular mechanisms

We cannot determine the penetrance of specific mutations observed here, as we do not see them often enough in an unselected population. Nevertheless, we introduce the term 'gene vulnerability' as the probability that a given type of mutation in a given gene contributes to a given condition. Genes with non-zero vulnerability define the vulnerable class. We can extend this concept to 'class vulnerability', defined as the mean gene vulnerability over a class of genes. Mathematically, class vulnerability, V , is computed by solving the following equation for V : $F \times A = P \times H \times V$, in which F is the prevalence of the condition, A is the ascertainment differential for DN mutations of a given type in the gene class, P is the expected proportion of the population with DN mutations of that given type, and H is the probability that such mutations hit the gene class.

We can compute a distribution of class vulnerability for all vulnerable genes targeted by a given mutational type (Methods) because F , A and P have empirically sampled distributions and H has a distribution inferred from the total length of the gene class. The distribution of class vulnerability for DN LGDs in males with lower IQ has a mode around 0.4 (Fig. 3). In other words, ~40% of DN LGDs in vulnerable genes in a male contribute to diagnoses of lower IQ ASD. Similarly, ~10% of missense mutations in vulnerable genes contribute to diagnoses of lower

IQ autism (Fig. 3). The mode for LGD vulnerability in females is fourfold lower than for lower IQ males, mainly because the prevalence is fourfold lower. Reduced penetrance in females is not well understood, but may be consequent to sexually dimorphic development. Support for this is seen in the relative enrichment of embryonically expressed genes as targets in females.

Partial gene vulnerability can be explained in several ways: some LGD mutations result in autism, some have little effect, and some produce other diagnoses or even lethality. Regardless, many LGD mutations will strongly predispose to ASD. We expect this to be reflected in decreased functional variation in the human gene pool, as we have previously shown for FMRP-associated genes⁸.

Given our analysis of gene vulnerability and the lack of evidence for compound heterozygosity, damage to a single allele will often have severe consequences for development. What underlies the vulnerability to haploinsufficiency is unclear. Half the normal gene dosage can result in half the level of gene products, and there are many examples in which physiology requires proper dosage^{32–37}. Also, having two copies of a gene will reduce variability of expression³⁸. With only one functional allele, there could be increased variation in levels of expression, including dangerously low levels at crucial moments in lineage development, altering the composition of tissues. Monoallelic expression also needs to be considered³⁹. Finally, some truncation events might lead to dominant negative alleles.

Present and future implications

From the clinical perspective, early diagnosis and family counselling are complicated if there are hundreds of genetic targets, especially if few are known with certainty. Sequencing of more cohorts is thus clearly warranted. From the therapeutic perspective, the good news is that in almost all cases DN mutations occur in probands in whom a normal allele is also present. It is theoretically possible that enhancing activity of the remaining alleles might alleviate symptoms. So in our view, the long-term prognosis for treating ASD is positive. Moreover, ASD targets overlap with targets for intellectual disability and schizophrenia, so mechanism-based treatments might work for different diagnostic categories. In the intermediate term, functional clustering suggests that treatments might be tailored to a smaller number of convergent pathways.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 July; accepted 3 October 2014.

Published online 29 October 2014.

- Jeste, S. S. & Geschwind, D. H. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nature Rev. Neurol.* **10**, 74–81 (2014).
- Sanders, S. J. *et al.* Multiple recurrent *de novo* CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).
- Levy, D. *et al.* Rare *de novo* and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70**, 886–897 (2011).
- Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
- Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
- Sanders, S. J. *et al.* *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
- O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
- Iossifov, I. *et al.* *De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
- Ronemus, M., Iossifov, I., Levy, D. & Wigler, M. The role of *de novo* mutations in the genetics of autism spectrum disorders. *Nature Rev. Genet.* **15**, 133–141 (2014).
- Zhao, X. *et al.* A unified genetic theory for sporadic and inherited autism. *Proc. Natl Acad. Sci. USA* **104**, 12831–12836 (2007).
- Fischbach, G. D. & Lord, C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010).
- Campbell, C. D. *et al.* Estimating the human mutation rate using autozygosity in a founder population. *Nature Genet.* **44**, 1277–1281 (2012).
- Michaelson, J. J. *et al.* Whole-genome sequencing in autism identifies hot spots for *de novo* germline mutation. *Cell* **151**, 1431–1442 (2012).

14. Schrider, D. R., Hourmozdi, J. N. & Hahn, M. W. Pervasive multinucleotide mutational events in eukaryotes. *Curr. Biol.* **21**, 1051–1054 (2011).
15. Kong, A. *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
16. Neale, B. M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
17. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
18. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
19. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
20. Bayés, A. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nature Neurosci.* **14**, 19–21 (2011).
21. Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E. & Eppig, J. T. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res.* **39**, D842–D848 (2011).
22. Feldman, I., Rzhetsky, A. & Vitkup, D. Network properties of genes harboring inherited disease mutations. *Proc. Natl Acad. Sci. USA* **105**, 4323–4328 (2008).
23. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
24. Newschaffer, C. J. *et al.* The epidemiology of autism spectrum disorders. *Annu. Rev. Public Health* **28**, 235–258 (2007).
25. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
26. Fromer, M. *et al.* *De novo* mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
27. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genet.* **45**, 984–994 (2013).
28. McCarthy, S. E. *et al.* *De novo* mutations in schizophrenia implicate chromatin remodeling and support a genetic overlap with autism and intellectual disability. *Mol. Psychiatry* **19**, 652–658 (2014).
29. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
30. O'Roak, B. J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622 (2012).
31. Nishiyama, M., Skoultschi, A. I. & Nakayama, K. I. Histone H1 recruitment by CHD8 is essential for suppression of the Wnt- β -catenin signaling pathway. *Mol. Cell. Biol.* **32**, 501–512 (2012).
32. Birchler, J. A. & Veitia, R. A. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl Acad. Sci. USA* **109**, 14746–14753 (2012).
33. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* **132**, 1077–1130 (2013).
34. Darnell, J. C. Defects in translational regulation contributing to human cognitive and behavioral disease. *Curr. Opin. Genet. Dev.* **21**, 465–473 (2011).
35. Veitia, R. A., Bottani, S. & Birchler, J. A. Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. *Trends Genet.* **29**, 385–393 (2013).
36. Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature Rev. Genet.* **14**, 125–138 (2013).
37. Zhang, F., Gu, W., Hurles, M. E. & Lupski, J. R. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* **10**, 451–481 (2009).
38. Eckersley-Maslin, M. A. & Spector, D. L. Random monoallelic expression: regulating gene expression one allele at a time. *Trends Genet.* **30**, 237–244 (2014).
39. Jeffries, A. R. *et al.* Random or stochastic monoallelic expressed genes are enriched for neurodevelopmental disorder candidate genes. *PLoS ONE* **8**, e85093 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements Simons Foundation Autism Research Initiative grants to E.E.E. (SF191889), M.W.S. (M144095 R11154) and M.W. (SF235988) supported this work. Additional support was provided by the Howard Hughes Medical Institute (International Student Research Fellowship to S.J.S.) and the Canadian Institutes of Health Research (Doctoral Foreign Study Award to A.J.W.). E.E.E. is an Investigator of the Howard Hughes Medical Institute. We thank all the families at the participating SSC sites, as well as the principal investigators (A. L. Beaudet, R. Bernier, J. Constantino, E. H. Cook Jr, E. Fombonne, D. Geschwind, D. E. Grice, A. Klin, D. H. Ledbetter, C. Lord, C. L. Martin, D. M. Martin, R. Maxim, J. Miles, O. Ousley, B. Peterson, J. Piggot, C. Saulnier, M. W. State, W. Stone, J. S. Sutcliffe, C. A. Walsh and E. Wijsman) and the coordinators and staff at the SSC sites for the recruitment and comprehensive assessment of simplex families; the SFARI staff for facilitating access to the SSC; and the Rutgers University Cell and DNA Repository (RUCDR) for accessing biomaterials. We would also like to thank the CSHL Woodbury Sequencing Center, the Genome Institute at the Washington University School of Medicine, and Yale Center for Genomic Analysis (in particular J. Overton) for generating sequencing data; E. Antoniou and E. Ghiban for their assistance in data production at CSHL; and T. Brooks-Boone, N. Wright-Davis and M. Wojciechowski for their help in administering the project at Yale. The NHLBI GO Exome Sequencing Project and its ongoing studies produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010).

Author Contributions CSHL: I.I., M.R. and M.W. designed the study; I.I., D.L., B.Y., Y.L., E.G., E.D., P.A., A.L., J.K., G.N., S.Y., M.C.S., K.Y. and M.W. analysed the data; M.R., I.H., J.R., B.M., L.R., J.T. and W.R.M. generated the exome data at Cold Spring Harbor Laboratory; I.I., Z.W., S.M. and J.T. confirmed the variants; I.I., M.R. and M.W. wrote the paper. UCSF/Yale: S.J.S. and M.W.S. designed the study; S.J.S., S.D., L.W. and A.J.W. analysed the data; S.J.S., J.D., L.E.G., J.D.M., C.A.S., M.F.W. and Z.W. confirmed the variants; S.M.M. and M.T.M. generated the exome data at Yale Medical Center. UW: B.J.O., J.S. and E.E.E. designed the study; B.J.O. and N.K. analysed the data; B.J.O., H.A.S., K.T.W. and L.V. confirmed the variants; E.E.E. and J.S. revised the manuscript; K.E.P., J.D.S., B.P. and D.A.N. generated the exome data at the University of Washington.

Author Information Sequence data used in these work are available from the National Database for Autism Research (<http://ndar.nih.gov/>), under study DOI:10.15154/1149697. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S. (shendure@uw.edu), E.E.E. (eee@gs.washington.edu), M.W.S. (matthew.state@ucsf.edu), or M.W. (wigler@cshl.edu).

METHODS

Sample collection. Most families (2,517) came from current or former members of the SSC. The SSC was assembled at 13 clinical centres, accompanied by detailed and standardized phenotypic analysis as reported previously¹¹. Several IQ measures (verbal, non-verbal and full spectrum) were recorded; in this work, we stratified probands by non-verbal IQ, which we refer to as simply 'IQ' throughout the text. Families with single probands and unaffected siblings were preferentially recruited, whereas families with two probands were specifically excluded¹¹. Families from two associated collections were also sequenced: the Simons Ancillary Collection (SAC, $n = 123$), and the Simons Twin Collection (STC, $n = 13$). The SAC includes families that failed inclusion criteria for the SSC, typically because a parent, sibling or second- or third-degree relative of the affected participant has been diagnosed with ASD, or for cases in which the proband's ASD diagnosis was questionable. The STC consists of families of monozygotic twins in which at least one co-twin is affected by ASD. The institutional review boards of Cold Spring Harbor Laboratory, Yale Medical Center and University of Washington, Seattle approved this study. Written informed consent from all subjects was obtained by SFARI. Blood samples were drawn from parents and children (affected and unaffected) and sent to the Rutgers University Cell and DNA Repository (RUCDR) for DNA preparation. DNAs from 2,517 families (of ~2,800 total in the SSC) were used in this study. Results from 774 of the SSC families included here were published in earlier work^{6–8}. The samples were split across the three centres: Cold Spring Harbor Laboratory (CSHL), the Department of Genetics at the Yale School of Medicine (YALE), and Department of Genome Sciences at the University of Washington (UW). The split was not uniform with respect to number of families or the proportions of female probands and probands with lower IQ (Extended Data Fig. 2 and Supplementary Table 1). Several families were sequenced at multiple centres, with 24 families sequenced in all three centres (Extended Data Fig. 1 and Supplementary Table 1).

Exome capture, sequencing and validation. The three centres differed in the precise exome capture platform, read length and validation protocols.

CSHL. The protocols described previously⁸ were applied to the families newly sequenced at CSHL. In brief, SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen) reagents were used with a custom barcoding protocol that enabled simultaneous exome enrichment of ≤ 4 genomes and the sequencing of ≤ 8 individuals per Illumina HiSeq 2000 lane. All exome sequencing was performed using paired-end 100-base pair (bp) reads. All strong and weak LGD candidate variants as well as additional variants from families sequenced at CSHL were subjected to experimental validation. Gene-specific primers were designed for PCR amplification of candidate single nucleotide variables (SNVs) and indels, and amplicons were pooled and sequenced on an Illumina MiSeq. Approximately 100 variants were validated per lane with paired-end 150-bp reads. Where possible, the parental origin was determined by phasing of linked transmitted SNVs.

UW. Samples were captured and sequenced by one of three methods. In the pilot set (19 quads), samples were captured using SeqCap EZ Human Exome Library v1.0 (Roche NimbleGen) reagents (UW-M1)^{7,40}. The remaining samples were captured using SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen) reagents⁷. Newly sequenced samples were either processed as described previously⁷ (UW-M2) or with a modified (UW-M3) protocol (Supplementary Table 10). For UW-M2, single-plex captures and single-plex sequencing runs (non-pooled) were performed as described previously⁷. For UW-M3, single-plex capture was performed as in UW-M2; however, in the post-capture PCR, an 8-bp index barcode was added. Post-PCR libraries were quantified and pooled in sets of ~96. These pools were then sequenced on the Illumina MiSeq platform to evaluate library complexity and sample distribution. Pools were rebalanced on the basis of performance, then sequenced across multiple HiSeq 2000 lanes using paired-end 50-bp reads. Additional lanes were added until samples reached target coverage (20 \times : ~80%; 8 \times : ~90%). If additional coverage was required for some samples, subpools were also generated. For samples processed with UW-M1 and UW-M2, predicted *de novo* calls were validated using standard PCR and Sanger sequencing⁷. For UW-M3 processed samples, custom Molecular Inversion Probe (MIP) capture probes were designed with targeting arms flanking regions of interest. Probes were designed without or with degenerate tags, and pools of ~50–100 probes were generated^{40,41}. As described earlier⁴¹, sets of families (~96 samples) were captured using these pools with 50–100 ng of genomic DNA as template. Capture products were then pooled and sequenced on an Illumina MiSeq. Candidate sites failing MIP QC or capture, or showing evidence of significant shifts in allele balance, were validated using the standard PCR/Sanger method. If sites repeatedly failed the assay, they were discarded. Novel sites called by the CSHL pipeline were validated using the same methods as UW-M3.

YALE. Whole blood-derived genomic DNA was enriched for exonic sequences using SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen) reagents. All family members were barcoded and each pool of four samples was sequenced using 75-bp paired-end reads on single lanes of the Illumina HiSeq 2000 instrument.

Where possible, all four family members were sequenced on the same lane to minimize batch effects. All strong and weak LGD candidate variants from the CSHL pipeline, along with an additional set of LGD candidates from the Yale pipeline, were subjected to experimental validation as follows: variant-specific primers were designed for PCR amplification of candidate SNVs and indels from all family members, and amplicons were sent for Sanger sequencing.

Sequence analysis pipelines. Sequence data were interpreted as family genotypes using pipeline tools at each respective data centre. Almost all of the data were reanalysed with the CSHL pipeline. We show the coverage (Extended Data Fig. 6) and yields (Extended Data Fig. 7) for *de novo* calls from each centre. The 24 families sequenced at all three centres demonstrated good agreement between pipelines and platforms (Supplementary Tables 11 and 12).

The analysis pipelines generated candidate *de novo* events, defined as variants present in the child and absent in both parents. We filtered out variants seen frequently in the parents of the collection (allele frequency >0.3%), reasoning that most of these would be false positives owing to uneven coverage in a parent. Candidates generated by local pipelines or by the common CSHL pipeline were validated at the respective centres with re-sequencing and 2,504 were verified. In our final call set, we include all verified calls from each centre, and omit any call that was rejected. In addition, because almost all (1,640 of 1,644) strong point mutations generated by the common CSHL pipeline were verified when successfully tested (Supplementary Table 11), such strong candidates are included in our call set even if the validation test failed or if the candidate event was not tested. All frameshift mutations were validated, and we exclude all that were rejected. All *de novo* calls used in the subsequent analysis, along with their validation status, are listed in Supplementary Table 2. Pipelines for analysis and validation were blind with respect to affected and unaffected status.

CSHL (uniform) pipeline. Sequence data from the three centres were analysed with the computational pipeline described previously⁸. In brief, the Illumina analysis pipeline (CASAVA 1.8) was used for base calls, and custom software was used to de-multiplex reads and trim barcodes from CSHL derived data. Data from Yale and UW were de-multiplexed at the respective centres before analysis through the CSHL pipeline. BWA⁴² was used to align sequence reads to the hg19 reference genome, and both Picard (<http://broadinstitute.github.io/picard/>) and GATK⁴³ were used for marking PCR duplicates, family-based sequence realignment and quality score recalibration. As described previously, a multinomial model-based family genotyper was used to generate candidate SNV and indel 'Mendel violators,' each annotated with: (1) a confidence score (denovoScr) that reflects the posterior probability of Mendel violation at the locus; (2) a goodness-of-fit score (chi2Score) showing the degree to which the assumptions of the multinomial model are applicable to the observed data; (3) counts of reads per allele and per family member; and (4) allele frequency and noise rates for the candidate position based on the whole collection. Candidates SNVs with denovoScr ≥ 60 and chi2Score >0.0001 were labelled 'strong' provided that the position was not polymorphic or noisy in the population, and that the parents were homozygous for the reference allele.

For SNVs, a cut-off denovoScr value of 60 was dictated by the desire to keep false positives to a minimum, and was chosen after computing the proportion of *de novo* candidates that appear at polymorphic loci (a surrogate for false positives) as a function of the score (see the Supplement of ref. 8). The low false positive rate (<5%) was also confirmed through experimental validation (Supplementary Table 11). In addition, we observe that only 1% of DN mutations are shared between two siblings (Supplementary Table 2), putting a 3% cap on false positives owing to failure to observe parents correctly. At stringent thresholds the false negative rate is generally high, but through simulations we determined that, even with stringent thresholds, regions with deep coverage (40 \times or higher joint coverage) had low false negative rates (<5%).

Indels were treated differently than SNVs. The multinomial model assumes a small allele bias, appropriate when calling SNVs, but not for *de novo* indels—particularly for long events (>10 bp). To address this, cut-offs for 'strong' indels were lowered (denovoScr >30 and chi2Score >10⁻⁹). To reduce noise, we added requirements for 'clean' read counts: parents were not allowed to have any reads containing the candidate indel, and were required to have at least 15 reads supporting the reference allele. At least one of the children had to have ≥ 6 reads with the candidate variant, and those reads had to comprise $\geq 5\%$ of reads. Experimental validation demonstrated that the false positive rate in the strong indels is <10%, and simulations for indels without extreme allele bias (most of those <10 bp) reveal that the false negative rate in well-covered regions (40 \times) is <5%.

All 'strong' SNVs and indels are reported here unless rejected by validation. To address the high false negative rates, we defined a class of 'weak' SNV and indels drawn from thresholds lower than strong candidates. All weak LGD candidates were subjected to validation, and only those successfully validated are reported. In addition, during method development (for example, through manual inspection or see ref. 44), we validated a large number of candidates that did not meet even

the weak definition. Candidate variants found as valid under these circumstances are reported here and labelled as 'not called'. This label is also used when the CSHL uniform pipeline missed a call from the UW and YALE data that was successfully validated.

UW pipeline. All samples using UW-M1 and UW-M2 protocols were processed as described earlier⁷. For UW-M3, updated versions of BWA (0.5.9-r16), Picard-tools (1.48) and GATK (1.0-6125) were used. The GATK Unified Genotyper was used in single sample mode with filter flags ($AB > 0.75$, low quality, $QD < 5.0$, $QUAL \leq 50.0$) and in parallel with the SAMtools pipeline as described previously⁴⁰ (see the GATK Unified Genotyper for definitions of these scores). Only positions with ≥ 8 -fold coverage were considered. Child genotype calls were compared to the parental genotypes to identify possible *de novo* events. Predicted DN SNVs were analysed against a set of 946 exomes to remove recurrent artefacts and likely undercalled sites. Indels were also called with the GATK Unified Genotyper and SAMtools⁴⁵, and included only those with $\geq 25\%$ of reads showing a variant at a minimum depth of $8\times$. These were then filtered against a larger set of 1,779 exomes (as with SNVs). Those sites passing (that is, not present) in the exome screen and also not present in multiple UW-M3 processed families were manually evaluated by inspecting alignments in the Integrative Genomics Viewer (<http://www.broadinstitute.org/igv/home>). Sites with obvious misalignments (for example, non-gapped indels or soft-clipped only reads) were removed. Moreover, if reads supporting the predicted DN mutation were present in $\geq 5\%$ of 20 (or more) reads in one of the parents, the site was excluded. For sites with lower coverage, a variant was excluded if present in $\geq 10\%$ (for example, 1 in 10, or 2 in 20) of parent reads or (for quads) if at least one variant read was present in one parent and the other child.

Yale pipeline. The Yale data were analysed as described previously⁶. In brief, CASAVA 1.8 was used for demultiplexing and base calling, reads were aligned to hg19 with BWA⁴², and SAMtools⁴⁵ was used for marking PCR duplicates and genotyping. In-house scripts were used for family-based assessment of *de novo* mutations and annotation against genes and the exome variant server (<http://varianttools.sourceforge.net/Annotation/EVS>).

Null models for target overlaps and recurrence. We introduce the term mutation-child-type to refer to a set of events of a certain mutational type (for example, missense or LGD) in children of a certain type (for example, male affected individuals with higher IQ or unaffected siblings). We observe target enrichment in gene classes, and document overlaps and recurrence between and within mutation-child-types. To measure significance, we use a null model in which the probability that a gene is hit by mutation is proportional to its length, a model supported by observation (Extended Data Fig. 5). We examine the distributions of lengths of gene targets of *de novo* synonymous, missense and LGD mutation in affected children and siblings. These distributions are compared to simulations of genes picked at random or in proportion to their length. The data fit well with the model that mutation frequency is linearly dependent on gene length. The group with the largest deviation from this rule is the set of DN targets in affected children, both for missense ($P = 0.001$) and for LGDs ($P = 0.001$, Supplementary Table 13). These P values are defined as the probability that the median length of the target class can arise under the null model, and are computed by simulations of equal number of genes weighted by length. Although the deviation is statistically significant, it is of such a minor amount that we ignore it for the null model.

Measuring overlaps. We test for overlaps between targets of a given mutation-child-type and other sets of genes (for example, overlap of DN LGD targets in affected girls with FMRP-associated genes) as well as overlaps between targets of two different mutation-child-types (for example, overlap between the targets of DN missense in all probands and the targets of DN LGDs in all probands). In both cases, observed overlaps are compared to those expected under the length-based null model discussed above.

Let T be the targets of mutation of a given type in a child of a given type, S a predefined gene set, and O the intersection of T and S . We ask for any gene G that carries a single mutation, what the probability $p(S)$ is that the mutation (and hence G) falls in S . We estimate $p(S)$ by collapsing all recurrent hits to one, and applying the length-based null model to S . Thus $p(S)$ is the ratio of (1), the sum of exome-captured lengths of the genes of S , divided by (2), the sum of the exome-captured lengths of all genes. Supplementary Table 7 shows the length of the captured portion of all genes in the exome we analyse. Using $|\cdot|$ to designate the number of members in a set, we then perform a two-sided binomial test of $|O|$ outcomes in $|T|$ opportunities given the probability of success $p(S)$.

When we test overlaps between targets of two different mutation-child-types, we take one of the targets as T and compare the other targets as S . However, before constructing T and S , we cleanup targets shared by T and S that result from mutations shared between siblings in the same family, or from multiple mutations of different types affecting a single gene in one child. We then apply the method of the paragraph above. Finally, we reverse the procedure for creating of S and T , and report both results (Supplementary Table 6).

Test for excessive recurrence. If we have R recurrent genes in K events in a mutation-child-type class, we test for excess recurrence by comparing R to the number of recurrent genes expected under the gene length-based null model. We build the expectation by performing 10,000 simulations. In each simulation, we sample K genes with replacement where the probability of sampling a gene is proportional to its length. We then count the number of recurrences.

Estimation of the number of vulnerable genes. To estimate the number of vulnerable genes for a given mutation-child-type, we start with the observed number of events (K), the observed number of recurrent events (R), the estimated posterior distributions for the rates of mutations of the given type in the ascertained (Mdist) and for the unaffected (Pdist) population. We then explore possible number of vulnerable genes (T) from 1 to 4,000. For each T , we estimate (through a simulation described in the next paragraph) the likelihood $L(T) = P(R|T, \text{Mdist}, \text{Pdist}, K)$. Assuming all numbers of vulnerable genes from 1 to 4,000 are equally likely, we compute a posterior distribution of the number of genes $p(T)$ proportional to $L(T)$ and determine the maximum value and 95% confidence intervals.

To estimate the likelihood, $L(T) = P(R|T, \text{Mdist}, \text{Pdist}, K)$, we perform 10,000 simulations for every T . In each simulation:

- (1) We randomly select T distinct vulnerable genes from all genes, without respect to length. Unlike mutation, which strikes a gene according to its length, we assume that the chance a gene can cause autism if mutated is independent of its length.
- (2) We select the number N of contributory events by sampling from a binomial distribution $\text{Binom}(K, A/M)$, where P a randomly selected rate from Pdist, M is a randomly selected from Mdist, $A = M - P$ is a sampled ascertainment differential, and A/M is an estimate of the proportion of contributory events.
- (3) We simulate N contributory mutation events by selecting N events with replacement from the T vulnerable genes proportional to their length.
- (4) To simulate random events, we select $K - N$ genes from all well-covered genes with replacement proportional to their length.
- (5) We record the number of recurrent events in the K selected events from above.

We set $L(T) = P(R|T, \text{Mdist}, \text{Pdist}, K)$ to be the proportion of simulations in which the number of recurrent events is exactly R . $P(T)$ is obtained by normalizing $L(T)$. For every simulation in which the number of recurrent events is exactly R , we also record the proportion of contributory events among the recurrent events, and the vulnerability point estimate as discussed in the next section.

Vulnerability. We use the equation described in the text: $F \times A = P \times H \times V$, in which F is the prevalence of the given condition in the population, A is the ascertainment differential for DN mutations of a given type in persons ascertained for that condition, P is the expected proportion of the population with such DN mutations, H is the probability that such a mutation hits the target, and V is the mean class vulnerability. These variables are in fact random variables with empirically derived distributions.

We first demonstrate the method for computing the class vulnerability point estimate for genes vulnerable to LGD mutations for the ASD males of lower IQ, assuming that the variables are fixed. One in 75 males is diagnosed with autism, and we estimate (from empirically derived gender biases) that three-quarters of these males are of lower IQ, yielding a prevalence $F = 1/100$. From our study, 0.23 of these have an LGD. Because the expected proportion of people with a DN LGD mutation is $P = 0.11$, only $A = 0.12$ of this subpopulation have an LGD in a vulnerable gene that contributes to ascertainment. Thus $F \times A = 1.2 \times 10^{-3}$ is the proportion of males that have lower IQ and autism resulting at least partially from a DN LGD. This proportion is also given by $P \times H \times V$, in which H is the probability that the LGD hits within the genes vulnerable to LGD mutations, and V is the mean class vulnerability for these genes. P , as already stated, is 0.11. We have computed the number of genes vulnerable to LGD mutations, N , for the affected males with lower IQ to be about 400 genes (Supplementary Table 6). Assuming membership in the target class is independent of gene length, and about 20,000 genes, we calculate $H = 400/20,000 = 0.02$, and solve V to be 0.55.

We assume the following prevalence: $F = 1/75$ for ASD in males, $F = 1/100$ for ASD with lower IQ in males, $F = 1/300$ for ASD with higher IQ in males, and $F = 1/300$ for ASD in girls. A and P are empirically derived gamma distributions from the sampled Poisson rates of DN LGD mutations in affected and unaffected siblings. By keeping the observed number of LGD events and the observed proportion of LGD events constant, we sample from the distribution of target number N and the distributions on A and P as described in the previous section. We set H to be the ratio of the total length of uniformly sampled vulnerable genes to the total length of the analysed captured exome, and compute a vulnerability point estimate as described just above. These sampled values are displayed in Fig. 3, bottom. The mode for V is 0.4 for males of lower IQ.

Parental age and phasing of DN mutations. We used two different strategies for modelling the relationship between rates of DN substitutions and the ages of the parents.

The first strategy does not depend on knowledge of the parent of origin for DN substitutions, which we do not know for the vast majority of DN substitutions. Because the ages of the mother and the father are strongly correlated, we can effectively use this strategy only to explore the relationship between the father's age and the rates of DN substitutions. Over probands and siblings in the 40×-joint family target, we model the number of mutations per child as sampled from a Poisson distribution with rate $R_c = T_c \times (A \times F_c + B)$, in which R_c is the rate of DN substitutions per child, F_c is the age of the father at the birth of the child, T_c is the ratio of the length of the 40×-target in that child to the total exome length, and A and B are whole population parameters, estimated by maximizing the likelihood over all children.

The second strategy is applicable only to DN mutations for which we have successfully 'phased' the parent of origin by proximity to a linked polymorphism. For each parental gender, we separately perform a two-sided one-sample *t*-test to compare the parental ages of each phased DN mutation to the mean of parental ages in our population.

DN substitutions increase ~0.4 per paternal decade (Extended Data Fig. 4), consistent with previous studies¹⁵ and the increase in autism as a function of paternal age^{46,47}. Where we could determine parental phase, DN substitutions arose more frequently in the paternal (287) than in the maternal (80) background. Among phased DN events, the mean age at birth was 34.6 for the father and 32.0 years for the mother, whereas the respective mean ages were 33.2 and 31.1 years for fathers and mothers in the whole population ($P = 0.0001$ and 0.047 , respectively, that these differences arise by chance).

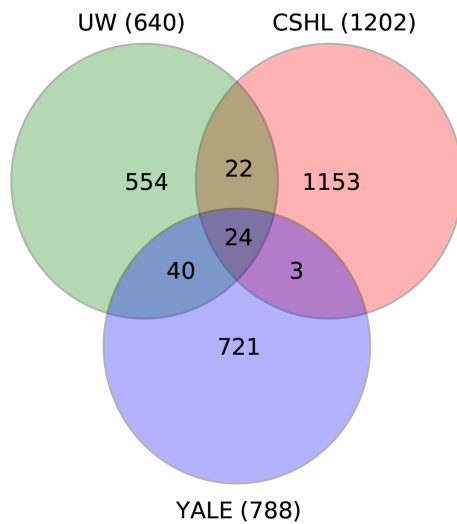
Gene class definition. For determining overlap with *de novo* mutations, functional gene classes were defined as follows. 'FMRP' are genes encoding transcripts that bind to FMRP¹⁷. 'Chromatin' indicates chromatin modifiers as defined by Gene Ontology (GO; <http://www.geneontology.org/>). 'PSD' is a set of genes encoding proteins that have been identified in postsynaptic densities²⁰. 'Mendelian' represent positionally identified human disease genes²², and 'essential' genes are human orthologues of mouse genes associated with lethality in the Mouse Genome Database²¹. 'dn LGD (Scz)' are *de novo* LGDs in schizophrenia^{26,48,49} and 'dn LGD (ID)' are *de novo* LGDs in intellectual disability^{25,29}.

'Embryonic' genes are those expressed in post-mortem human embryonic brains¹⁹, derived from downloaded expression data¹⁸ (<http://www.brainspan.org/static/>

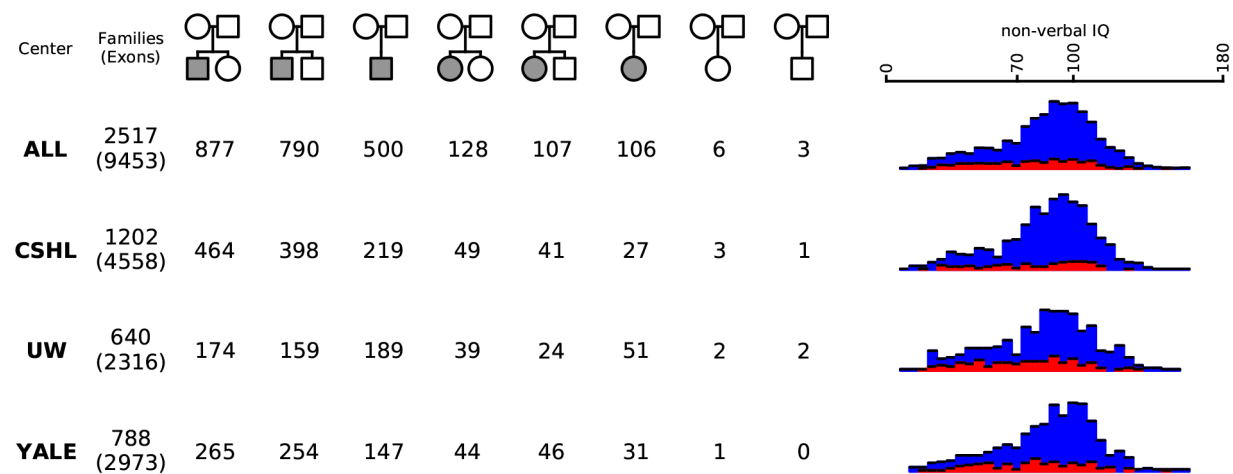
[download.html](#)). This data set provides normalized expression levels for ~17,000 genes across brain regions from 36 individuals, 18 of which were from embryos. Each brain was further subdivided into 14 anatomical regions for a total of 508 regions. We computed correlation values for the 17,000 genes, and generated a graph by connecting genes that had correlations >0.85, then identified connected components and averaged the expression of genes within these components as a function of the annotated age of the brain and by region. Each region is sorted first by age, then by type (Extended Data Fig. 8). The averaged normalized expression of the 1,912 genes in the first component decreases after birth, and hence we call this set embryonic.

Supplementary Table 7 shows the genes in the eight functional classes that are within the captured exome regions and were used in all analyses.

40. O'Roak, B. J. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. *Nature Genet.* **43**, 585–589 (2011).
41. Boyle, E. A., O'Roak, B. J., Martin, B. K., Kumar, A. & Shendure, J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* **30**, 2670–2672 (2014).
42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
43. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
44. Narzisi, G. *et al.* Accurate *de novo* and transmitted indel detection in exome-capture data using microassembly. *Nature Methods* **11**, 1033–1036 (2014).
45. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
46. Reichenberg, A. *et al.* Advancing paternal age and autism. *Arch. Gen. Psychiatry* **63**, 1026–1032 (2006).
47. Croen, L. A., Najjar, D. V., Fireman, B. & Grether, J. K. Maternal and paternal age and risk of autism spectrum disorders. *Arch. Pediatr. Adolesc. Med.* **161**, 334–340 (2007).
48. Gulsuner, S. *et al.* Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
49. Xu, B. *et al.* Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. *Nature Genet.* **43**, 864–868 (2011).

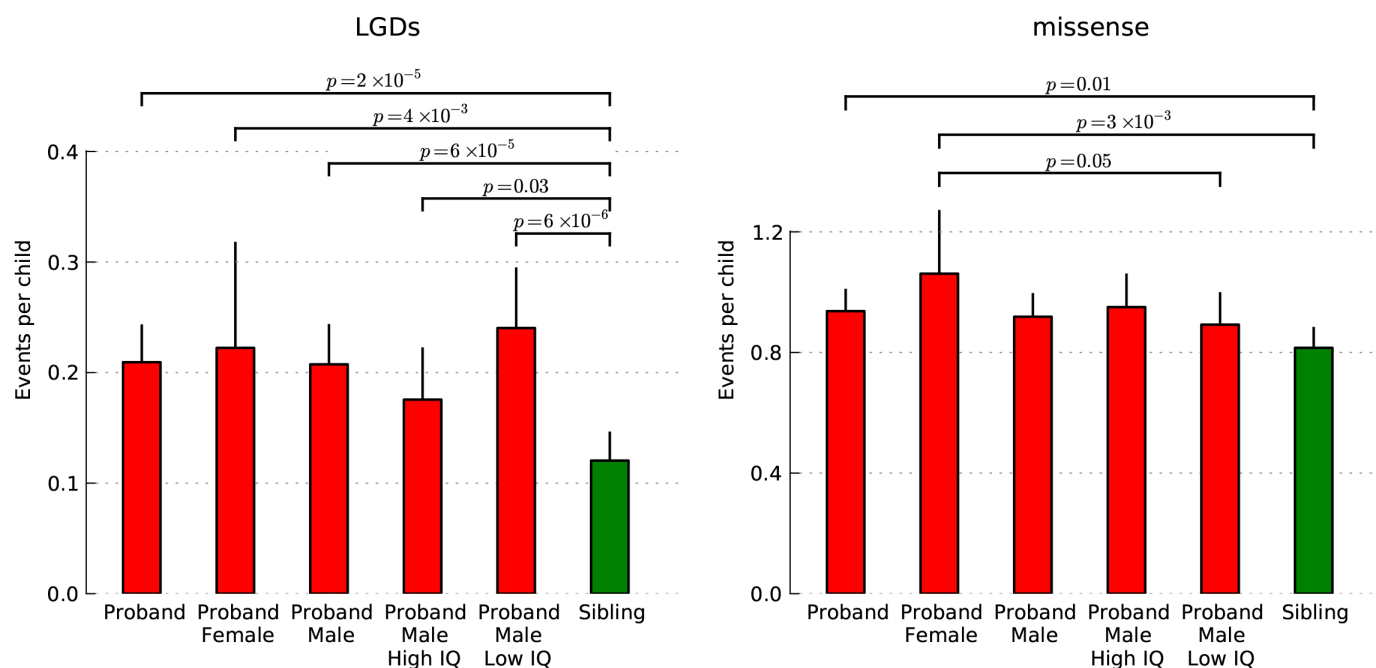


Extended Data Figure 1 | Number of families sequenced by centre. The numbers of families sequenced at the three centres are plotted as a Venn diagram. Families sequenced at more than one centre are indicated by the overlapping regions between circles. CSHL, Cold Spring Harbor Laboratory; UW, University of Washington, Seattle; YALE, Yale Medical Center.



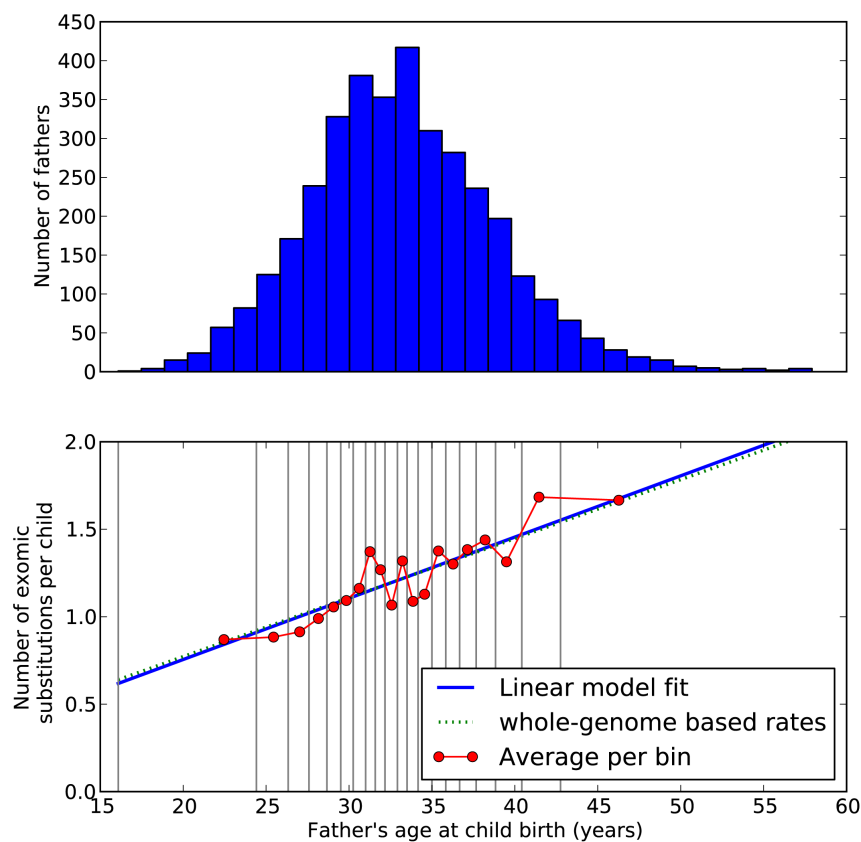
Extended Data Figure 2 | SSC sequencing by pedigree type and non-verbal IQ. A summary of all SSC families sequenced is indicated across the ‘all’ row. Numbers of SSC families with complete exome sequencing data are displayed by centre in the following rows (see Extended Data Fig. 1 legend for centre designations). The top number in entries under the ‘families’ column indicates the total number of families sequenced, and the number in parentheses below

indicates the total number of individuals. Family pedigree structures are shown across the top row with gender indicated by shape (square for male, circle for female) and affected status indicated by colour (white for unaffected, grey for affected). Distributions of non-verbal IQ within each cohort are shown for male probands (blue) and female probands (red).



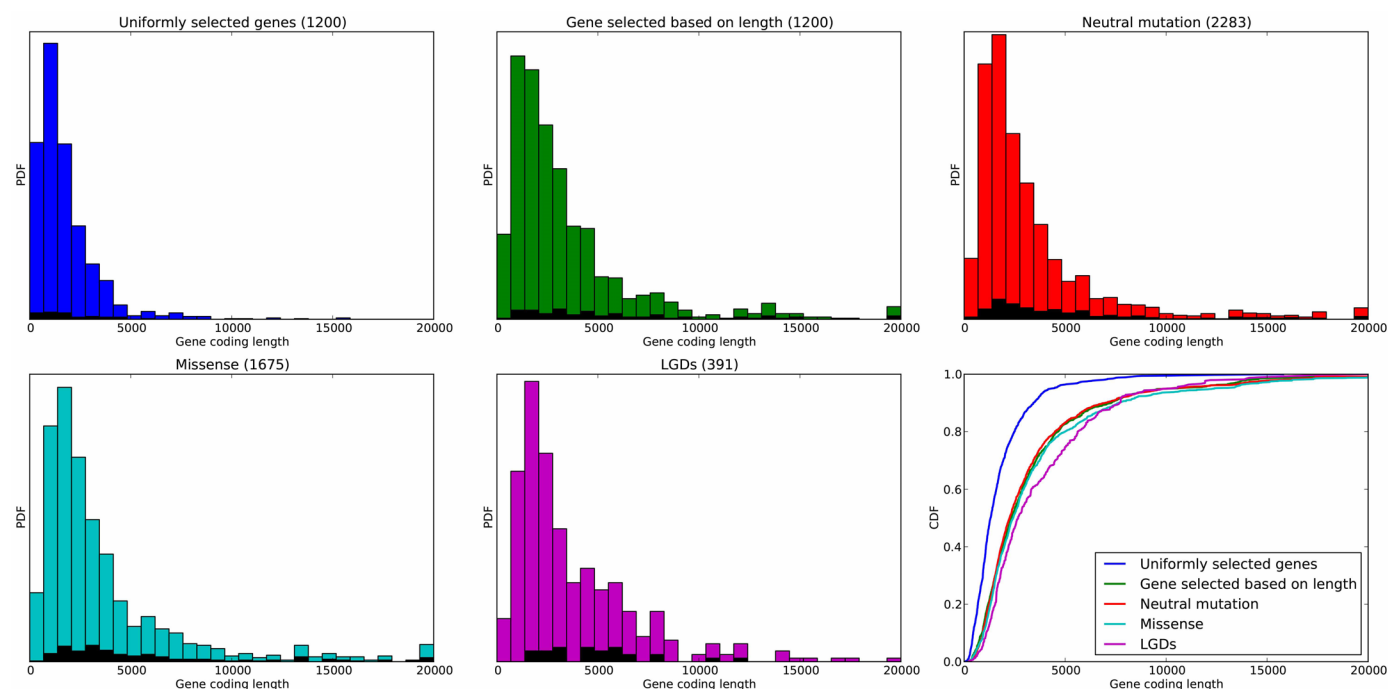
Extended Data Figure 3 | Rates of *de novo* LGD and missense mutations in the SSC by child status. On the left we show the LGD rate per child in six types of children, labelled on the *x* axis, defined by their affected status, gender, and non-verbal IQ. We test for equal rates for every pair of child types and we show the ones with $P > 0.05$ with thin lines on the top of the figure.

Although not significant, the rates in affected females and in affected males of lower non-verbal IQ are larger than the rate in males of higher non-verbal IQ. On the right, we show the missense rates per child for the same six groups of children.



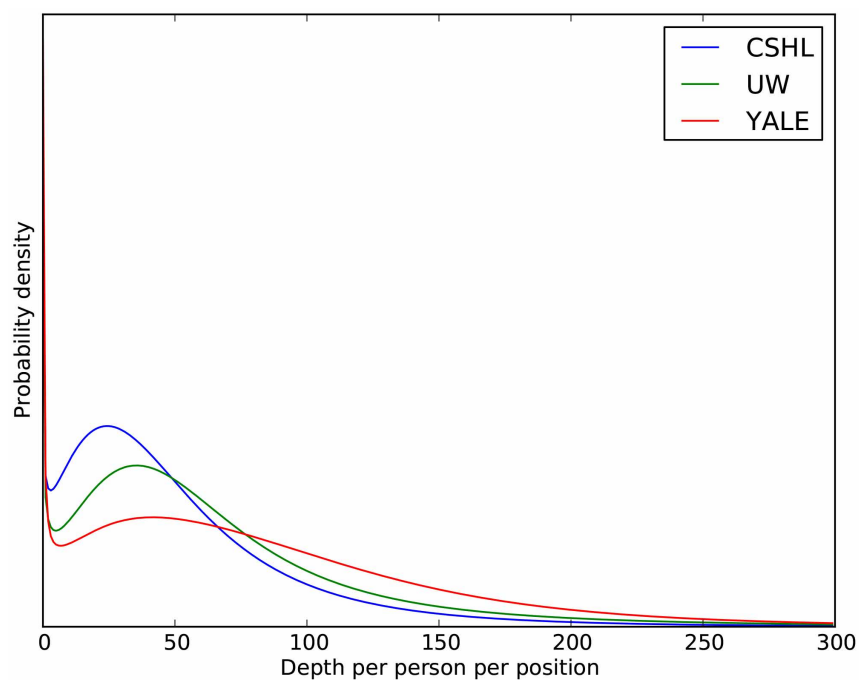
Extended Data Figure 4 | Paternal age and DN mutation rate at child birth. Distribution of paternal age at birth of children (top) and rates of DN mutation in offspring as a function of paternal age are shown (bottom). Children were ordered by paternal age at birth and split into 20 groups of similar size, as shown in the bottom panel. The red curve shows the mean observed rates of *de novo* exomic substitutions in each of the 20 groups, with the x coordinate

equal to the mean each of the fathers' ages within each group. The blue line shows a linear fit to the observed rates. The dotted green line represents DN mutation rates from whole genome sequencing data¹⁵ scaled to rates per exome based on representation in the SeqCap EZ Human Exome Library v2.0 (Roche NimbleGen).

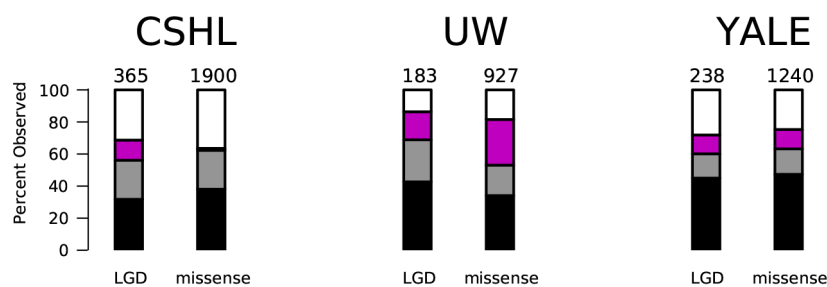


Extended Data Figure 5 | Coding region size distribution for query sets of genes. Probability density function (PDF) and cumulative distribution functions (CDF) (right bottom) of the distributions of the coding region length in base pairs of five sets of genes: a set of 1,200 genes picked uniformly from the set of exome-targeted genes (blue); a separate set of 1,200 genes picked with probabilities proportional to length of the coding region (green); the set of gene targets of neutral mutations, including synonymous mutations in probands and

siblings, and missense mutation in siblings (red); genes with *de novo* missense mutations in probands (cyan); and genes with *de novo* LGDs in probands (magenta). Black within the histograms shows the distribution of lengths of the recurrently hit genes from each class. Coding region length distribution under a uniform model does not fit the lengths of the genes with observed mutations, and genes with LGD mutations are longer than predicted by a simple length-based model (bottom right).

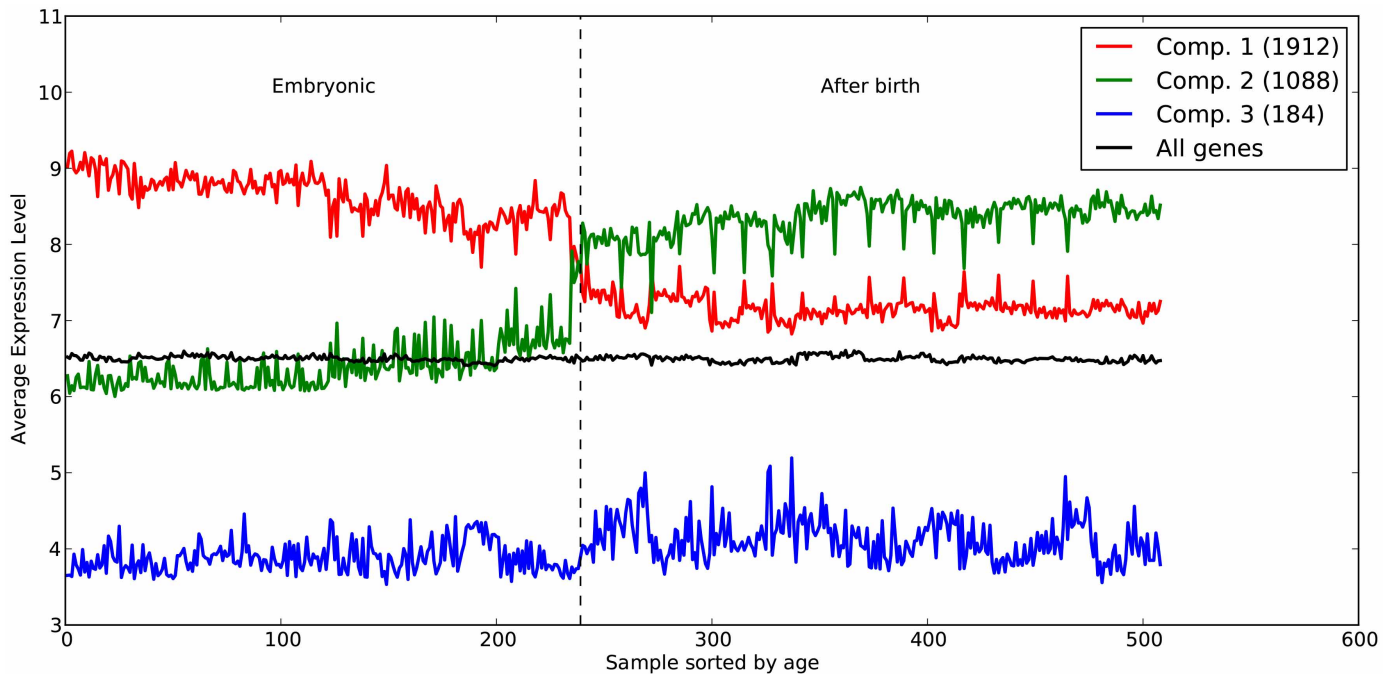


Extended Data Figure 6 | Distributions of sequencing depth. Distributions of sequencing depth (number of sequence reads covering a given genomic position) per person per position for the three sequencing centres are plotted. Centre designations are as in Extended Data Fig. 1.



Extended Data Figure 7 | Yield of DN LGD and missense mutations. We plot the yield of DN LGD and missense mutations per sequencing centre (designations as in Extended Data Fig. 1). In each case we show the number of mutations we expect to see based on the estimated rates per child, indicated by the numbers above the bars. We also show what percentage of the expected

number we have observed. Black refers to strong calls in the 40x target, grey refers to strong calls outside of 40x target, and magenta refers to weak (but valid) calls. The white region represents the difference between the expected and observed numbers of variants.



Extended Data Figure 8 | Categorization of embryonically expressed genes.

We downloaded expression data¹⁸ from <http://www.brainspan.org/static/download.html>. The data set provides normalized expression levels for ~17,000 genes across brain regions from 36 individuals, 18 of which were from embryos. Each brain was further subdivided into 14 anatomical regions for a total of 508 regions. We computed correlation values for the 17,000 genes, and generated a graph by connecting genes that had correlations >0.85 . We then

identified connected components and averaged the expression of genes within these components as a function of the annotated age of the brain and by region. Each region is sorted first by age, then by type. The averaged normalized expression of the 1,912 genes in the first component decreases after birth, and hence we call this set embryonic. See Supplementary Table 7 for the list of embryonic genes.

Evolution of mosquito preference for humans linked to an odorant receptor

Carolyn S. McBride^{1,2†}, Felix Baier^{1†}, Aman B. Omondi^{3†}, Sarabeth A. Spitzer^{1†}, Joel Lutomia⁴, Rosemary Sang⁴, Rickard Ignell³ & Leslie B. Vosshall^{1,2}

Female mosquitoes are major vectors of human disease and the most dangerous are those that preferentially bite humans. A 'domestic' form of the mosquito *Aedes aegypti* has evolved to specialize in biting humans and is the main worldwide vector of dengue, yellow fever, and chikungunya viruses. The domestic form coexists with an ancestral, 'forest' form that prefers to bite non-human animals and is found along the coast of Kenya. We collected the two forms, established laboratory colonies, and document striking divergence in preference for human versus non-human animal odour. We further show that the evolution of preference for human odour in domestic mosquitoes is tightly linked to increases in the expression and ligand-sensitivity of the odorant receptor *AaegOr4*, which we found recognizes a compound present at high levels in human odour. Our results provide a rare example of a gene contributing to behavioural evolution and provide insight into how disease-vectoring mosquitoes came to specialize on humans.

Blood-feeding as a behavioural adaptation is exceedingly rare in insects. Of the one million to ten million insect species on earth, only ~10,000 feed on the blood of live animals¹. Among these, only about 100 species blood-feed preferentially on humans¹. When biting insects evolve to prefer humans, they can spread diseases such as malaria and dengue

fever with devastating efficiency. The mosquito *Aedes aegypti* provides one of the best examples of specialization on humans. It originated as a wild, animal-biting species in the forested areas of sub-Saharan Africa, where the subspecies *Ae. aegypti formosus* is still often found living in forests and biting non-human animals today^{2–4}. In contrast, the derived

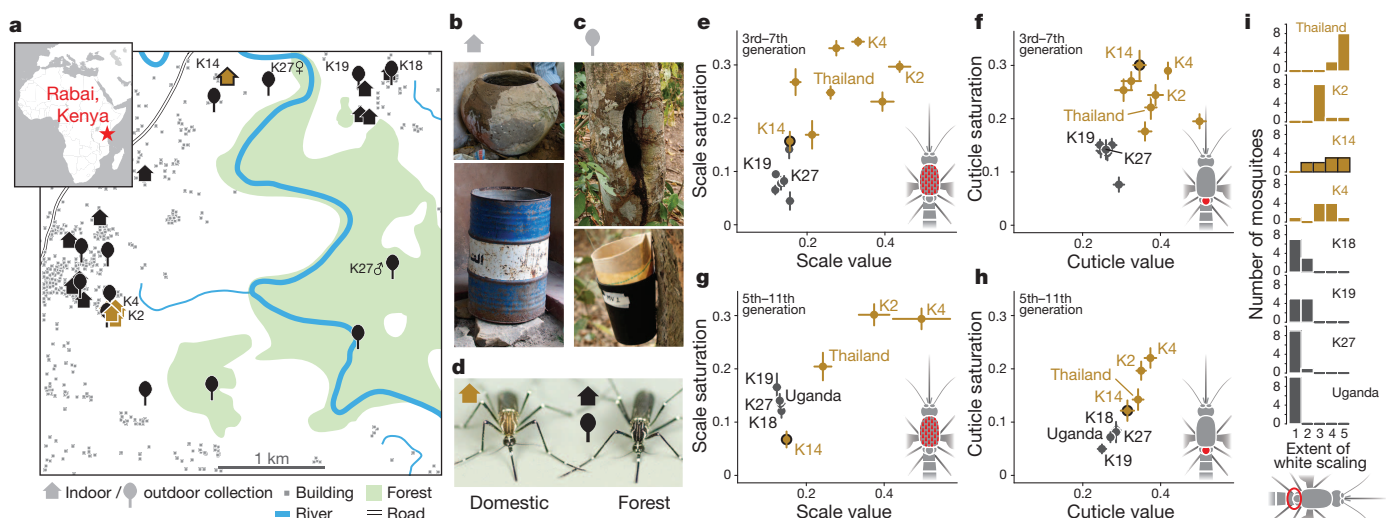


Figure 1 | Field collection of forest and domestic forms of *Ae. aegypti* in Rabai, Kenya. **a**, Map of Rabai, Kenya showing collection sites, with six colonies labelled. **b**, **c**, Typical indoor (**b**) and outdoor (**c**) water containers or traps. **d**, Examples of characteristic differences in body colour between domestic (left) and forest (right) females, selected from the extremes of variation. **e–h**, Summary of colour variation in scales (**e**, **g**) and cuticle (**f**, **h**) among laboratory colonies tested after 3–7 (**e**, **f**) and 5–11 (**g**, **h**),

including colonies from Thailand and Uganda. Cartoons at the bottom of **e–i** indicate in red where morphology was sampled. See also Extended Data Fig. 1. Data are plotted as mean \pm s.e.m. ($n = 10–15$ mosquitoes per colony). **i**, Extent of white scaling on the first abdominal tergite ($n = 10$ mosquitoes per colony). In all figures, black and brown indicate forest and domestic colonies, respectively, except K14 in brown with black outline to represent its mixed morphology.

¹Laboratory of Neurogenetics and Behavior, The Rockefeller University, New York, New York 10065, USA. ²Howard Hughes Medical Institute, 1230 York Avenue, New York, New York 10065, USA. ³Unit of Chemical Ecology, Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Box 102, Sundsvägen 14, 230 53 Alnarp, Sweden. ⁴Center for Virus Research, Kenya Medical Research Institute, PO Box 54840 – 00200, Off Mbagathi Way, Nairobi, Kenya. [†]Present addresses: Princeton Neuroscience Institute and Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA (C.S.M.); Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA (F.B.); Plant Protection Research Institute, Agricultural Research Council, Private Bag X134, Queenswood 0121, South Africa (A.B.O.); Harvard College, Harvard University, Cambridge, Massachusetts 02138, USA (S.A.S.).

non-African subspecies *Ae. aegypti aegypti* has evolved to specialize in biting humans and thus has become the main worldwide vector of dengue and yellow fevers^{2–4}.

The evolutionary adaptations that help subspecies *Ae. aegypti aegypti* exploit humans are most clearly seen where it has been reintroduced along the coast of East Africa and is known as the ‘domestic’ form. Researchers investigating the outbreak of an unknown illness in Tanganyika in 1952 discovered homes heavily populated by brown-pigmented ‘domestic’ mosquitoes⁵. Subsequent work in the Rabai region of Kenya in the 1960s and 1970s showed that domestic mosquitoes readily entered homes⁶, preferred to lay eggs in nutrient-poor river and rain water stored in containers indoors^{7,8}, were resistant to starvation as larvae⁹, and had evolved a strong preference for biting humans^{7,10,11}. Black-pigmented populations of the native African subspecies *formosus*, known in Rabai as the ‘forest’ form, were found just hundreds of metres away, avoiding homes, laying their eggs in tree holes and rock pools outdoors, and preferring to bite non-human animals. These differences translated into marked divergence in capacity to spread human diseases, including chikungunya, the unknown illness from 1952, yellow fever, prevalent in Africa and South America since the sixteenth century, and dengue fever, a disease currently infecting almost 400 million people around the world each year¹².

Remarkably, the domestic and forest forms in Rabai remained separate in nature but were interfertile in captivity⁸, providing a rare opportunity to investigate the genetic basis and evolution of traits that adapt mosquitoes to humans. Here we find that human host preference in domestic mosquitoes is strongly correlated with functional genetic variation in an odorant receptor, *Or4*, which recognizes a component of human body odour.

Domestic and forest forms continue to coexist

Forest and domestic mosquitoes were last documented in Rabai, Kenya in the 1970s, and we returned there in 2009 to determine whether they still exist. We collected *Ae. aegypti* larvae and pupae in water-storage containers inside approximately one in every five homes visited (Fig. 1a, b). We also collected eggs, larvae, and pupae of several mosquito species, including *Ae. aegypti*, outdoors in natural and artificial containers in villages and nearby forest (Fig. 1a, c). From these collections, we established 29 laboratory colonies, each descending from fewer than 20 males and females collected in the same house or outdoor location (Fig. 1a).

Previous reports described differences in body colour between the forms^{3,6}. Indeed, females from all outdoor colonies and some indoor colonies were black, resembling forest mosquitoes (Fig. 1d). Those from the remaining indoor colonies were brown, resembling domestic mosquitoes (Fig. 1d). Differences in thorax colour were maintained across multiple laboratory generations (Fig. 1e–h; Extended Data Fig. 1). Black and brown colonies also differed in abdominal scaling (Fig. 1i). A single indoor colony, K14, included individuals with a mix of black and brown traits (Fig. 1e–i). Black mosquitoes resembled a subspecies *formosus* colony from inland Africa (Uganda), whereas brown mosquitoes resembled a subspecies *aegypti* colony from Asia (Thailand) (Fig. 1e–i). In light of these morphological differences, and genetic differences among the field-collected progenitors of our colonies¹³, we hereafter refer to black and brown colonies as forest and domestic, respectively. In summary, mosquitoes fitting the morphological description of the two forms continue to coexist in Rabai, Kenya, 35 years after they were last documented.

Domestic mosquitoes strongly prefer humans

We used three assays to characterize the preference of forest and domestic forms for humans versus non-human animals. We offered guinea-pig as a non-human host because it is among the diverse hosts to which forest mosquitoes respond^{10,11}. In a biting assay in which females are exposed directly to live hosts (Fig. 2a), forest females preferred the guinea-pig (Fig. 2b) and domestic females weakly preferred the human (Fig. 2b). Domestic females were approximately twice as likely to respond overall (Fig. 2c), possibly reflecting adaptation to indoor environments, and by

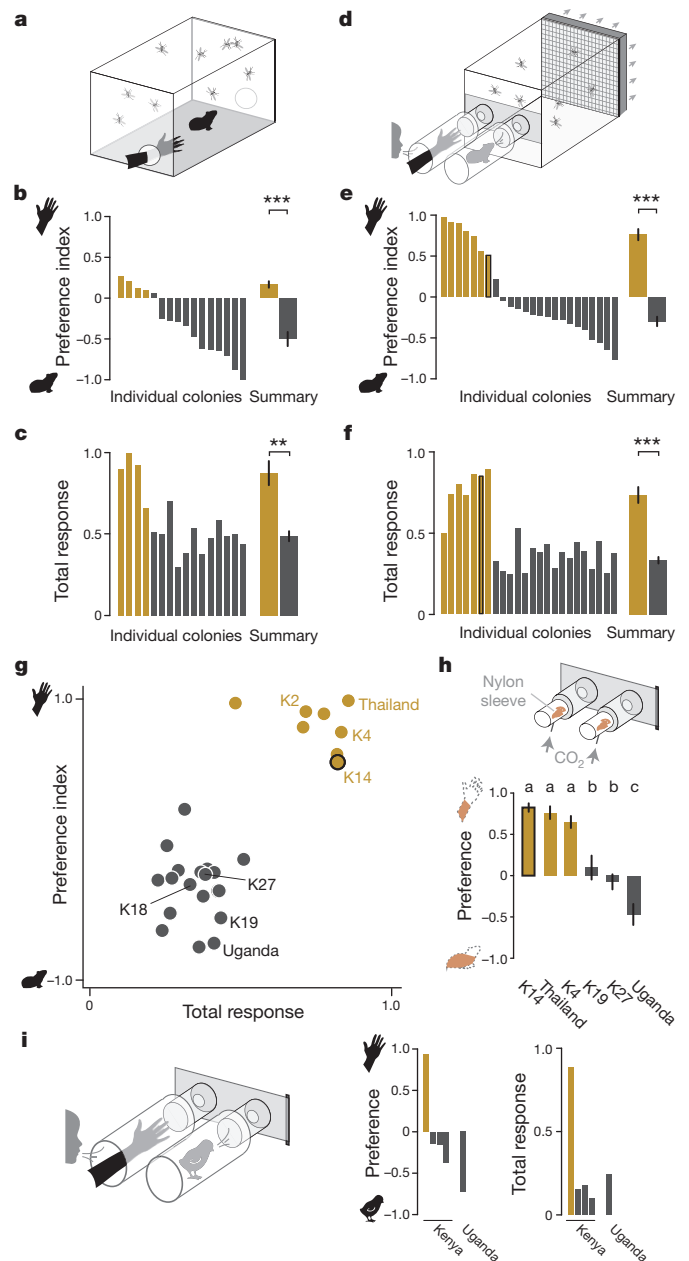


Figure 2 | Forest and domestic females differ in host preference. a–c, Biting assay. a, Assay schematic. b, c, Host preference (b) and total response of mosquitoes choosing either host (c) of 16 Kenyan colonies ($n = 1–3$ trials per colony). d–f, Live host olfactometer assay. d, Assay schematic. e, f, Host preference (e) and total response of mosquitoes choosing either host (f) of 24 Kenyan colonies ($n = 1–7$ trials per colony). Thin bars in b, c, e, and f are colony means and thick bars are summary mean of individual colony means \pm s.e.m. (two-sample *t*-test treats each colony as single data point). g, Summary of live host olfactometer behaviour of 24 Kenyan plus Thai and Ugandan colonies. h, Nylon sleeve olfactometer assay. Top, schematic of stimulus ports. Bottom, preference of 6 colonies (mean \pm s.e.m., $n = 5–17$ trials per colony). Bars labelled with different letters are significantly different (one-way ANOVA $P < 0.0001$, followed by Tukey’s test). i, Live host olfactometer assay (human versus chicken). Top, schematic of host ports. Bottom, preference and response of 5 colonies ($n = 1$ trial per colony). In all figures, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

extension laboratory settings¹⁴. When host cues were presented in an olfactometer (Fig. 2d), forest and domestic females again showed significantly different preferences and response rates (Fig. 2e, f), with domestic females displaying a strong preference for humans. All colonies fell into two discrete behavioural clusters corresponding precisely to the

forest and domestic designations made on the basis of morphology (Fig. 2g). Behaviourally, forest colonies resembled subspecies *formosus* from Uganda, and domestic colonies resembled subspecies *aegypti* from Thailand.

We further confirmed these results with host-scented nylon sleeves supplemented with equal amounts of carbon dioxide (CO₂), a potent activator of mosquito host-seeking¹⁵ (Fig. 2h). Three domestic colonies retained their strong preference for humans, whereas three forest colonies

ranged from no preference to moderate preference for guinea-pig (Fig. 2h).

To extend our results to other non-human hosts, we also assessed the preference of a small subset of colonies for human versus chicken in the live host assay, obtaining qualitatively similar results (Fig. 2i). Our findings confirm that domestic mosquitoes have evolved a marked preference for human body odour.

Human preference is associated with OR expression

Novel chemosensory preferences in insects are sometimes accompanied by changes in the peripheral chemosensory system^{16–20}. We reasoned that altered gene expression in antennae may contribute to preference, and profiled differential gene expression in this major olfactory organ using RNA sequencing (RNA-seq). To identify general differences between forms, we compared three forest versus three domestic colonies (Fig. 2h). To determine which of these differences are genetically associated with host preference²¹, we crossed two representative colonies and compared pools of strongly human-preferring versus guinea-pig-preferring F2 hybrids (Fig. 3a, b).

A total of 959 antennal genes were differentially expressed in colonies (Fig. 3c, e), 46 genes were differentially expressed in F2 pools (Fig. 3d, e), and 14 genes were differentially expressed in the same direction in both comparisons (Fig. 3c–e and Supplementary Table 1). Odorant receptors (ORs), a family of insect chemosensory receptors²² were significantly over-represented among differentially expressed genes (Fig. 3f, $P < 0.0001$). Two other families of chemosensory genes, the ionotropic receptors

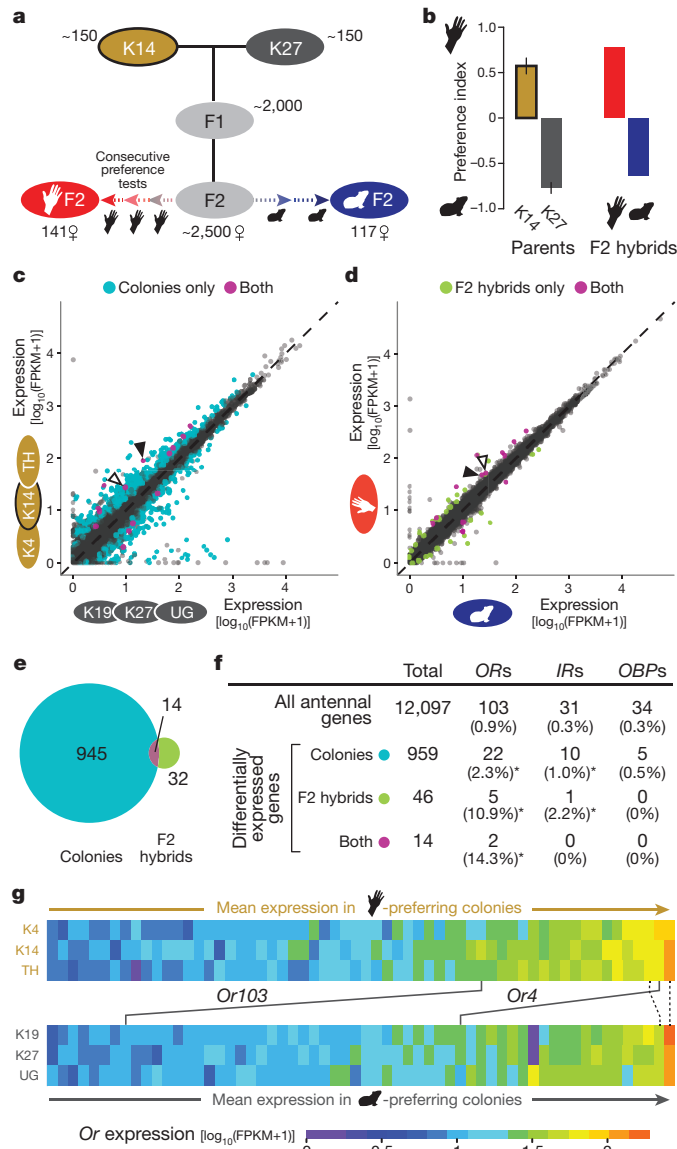


Figure 3 | Antennal gene expression is significantly associated with preference for humans. **a**, Schematic of isolation of strongly human-preferring and guinea-pig-preferring F2 females. **b**, Live host olfactometer preference of parent colonies ($n = 4-7$ trials, mean \pm s.e.m.) and F2 hybrids ($n = 1$). **c, d**, Differential antennal gene expression in colonies (**c**) and F2 hybrids (**d**) measured via RNA-seq (false discovery rate (FDR) = 0.05). Filled and open arrowheads point to *Or4* and *Or103*, respectively. **e**, Summary of differential gene expression in **c, d**. **f**, Numbers of detectable (fragments per kilobase of transcript per million mapped reads (FPKM) ≥ 1) (top row) or differentially expressed (bottom 3 rows) olfactory receptors (ORs), ionotropic receptors (IRs), and odorant-binding proteins (OBPs). Number as a percentage of total genes is shown in parentheses. Asterisks indicate significant enrichment relative to percentage of all antennal genes ($\chi^2 P < 0.0001$). **g**, Heat maps of ligand-selective OR genes with log₁₀FPKM expression > 0.83 in domestic (top) or > 0.84 in forest (bottom) colonies, with two ORs differentially (solid lines) or not differentially expressed (dashed lines) indicated.

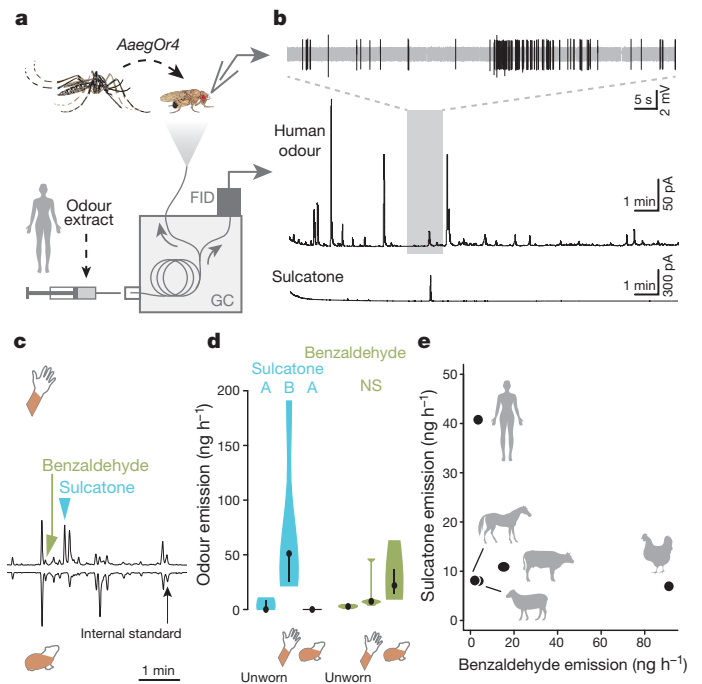


Figure 4 | *Or4* responds to sulcatone, a human odorant. **a**, Schematic of gas chromatography-coupled single sensillum recording (GC-SSR). FID, flame ionization detector. **b**, Sample GC-SSR output, with SSR trace at the top showing large-amplitude spikes (black) from the *Or4*-expressing neuron and FID output showing elution times of human volatiles (middle trace) and sulcatone (bottom trace). **c**, Magnified sections of sample total ion chromatogram traces from GC-mass spectrometry (GC-MS) analysis of human (top) and guinea-pig (bottom) sleeves. **d**, Summary of sulcatone and benzaldehyde emission from unworn, human, and guinea-pig sleeves ($n = 4-5$ replicates per group; violin plots clipped at the range of data; median: black dot; first and third quartile: bounds of black bar). Violins labelled with different letters are statistically different (pairwise Wilcoxon tests with Bonferroni correction; NS, not significant). **e**, Summary of sulcatone and benzaldehyde levels in human and animal odour collections ($n = 1$ collection from 1–20 individuals per species).

(IRs)²³ and odorant-binding proteins (OBPs)²⁴ were less enriched or not enriched, respectively (Fig. 3f). A selective role for the OR pathway in helping mosquitoes distinguish among hosts is consistent with previous work in a laboratory strain of *Ae. aegypti*²⁵.

Of the 14 genes significant in both colony and F2 comparisons, two were ORs (Fig. 3f), *Or4* and *Or103*. Both were upregulated in human-preferring mosquitoes (Fig. 3g, solid lines). *Or4* was also the second most highly expressed ligand-selective OR in the antennae of domestic females overall (Fig. 3g), and we chose this gene for further study.

Or4 recognizes the human odorant sulcatone

An olfactory receptor could modulate host preference by mediating attraction or repulsion to specific host odours, so we asked whether *Or4* is activated by a component of human or guinea-pig odour. We expressed the genome reference allele of *Or4* heterologously in a *Drosophila* olfactory neuron lacking a ligand-selective OR²⁶, and tested responses to fractionated host odour from guinea-pigs and humans (Fig. 4a). *Or4*

did not respond to any fraction of guinea-pig odour (data not shown), but responded consistently to a fraction of human odour corresponding to 6-methyl-5-hepten-2-one, commonly called sulcatone (Fig. 4b).

Sulcatone is a volatile odorant repeatedly identified in human body odour^{27–30}. Although sulcatone is also emitted by a variety of other animals^{31–34} and plants^{35–37}, it appears to reach uniquely high levels in humans (Fig. 4c–e). It was abundant in the odour of nylon sleeves worn by five humans, but undetectable or at low concentration in unworn sleeves or sleeves worn by four guinea-pigs (Fig. 4c, d). For reference, another widespread volatile, benzaldehyde, did not differ significantly between samples (Fig. 4c, d). We also found approximately four times more sulcatone in the body odour of live humans than in the odour of a live chicken or the hair of horses, cows, and sheep (Fig. 4e). Collectively, these results suggest that increased expression of *Or4* may help mosquitoes distinguish humans from non-human animals by conferring sensitivity to sulcatone. Interestingly, the malaria mosquito *Anopheles gambiae* has at least 4 ORs that are strongly activated by sulcatone³⁸, but none is closely related to *AaegOr4* (ref. 39). These two species diverged from each other approximately 150 million years ago and evolved independently to specialize in biting humans⁴⁰.

Preference linked to *Or4* sensitivity and expression

Evolution of preference for human hosts could occur not only via changes in *Or4* expression, but also via changes in the *Or4* coding region that affect protein function. Natural variation in chemoreceptor proteins has previously been shown to alter ligand-sensitivity⁴¹ and odour perception⁴². We found extensive variation in *Or4*, with seven major alleles present in the two parent colonies and F2 hybrids (Fig. 5a–c; Extended Data Fig. 3). The domestic parent, K14, was dominated by the closely related A and B alleles (Fig. 5a, b) and a highly divergent G allele (Fig. 5a, b). The K27 forest parent, in contrast, harboured 5 distinct alleles at low to moderate frequency (Fig. 5a, b). RNA-seq data from 8 additional colonies suggest that these patterns apply globally. Human-preferring colonies derived from Kenya, Thailand, USA, and West Africa were dominated by A-like alleles, whereas animal-preferring colonies from Kenya and Uganda were all highly variable (Extended Data Fig. 2). Although all alleles were present in F2 mosquitoes, they were inherited at different frequencies by human-preferring and guinea-pig-preferring individuals (Fig. 5c, d). Moreover, F2s tended to carry alleles characteristic of the parent with similar preference (compare Fig. 5b and 5d), suggesting that some aspect of allele-specific function affects preference.

We asked how allelic variation affects *Or4* receptor function. The protein-coding sequence of *Or4* is remarkably variable among alleles with differences in 13 of 406 residues on average, and 26 residues in the

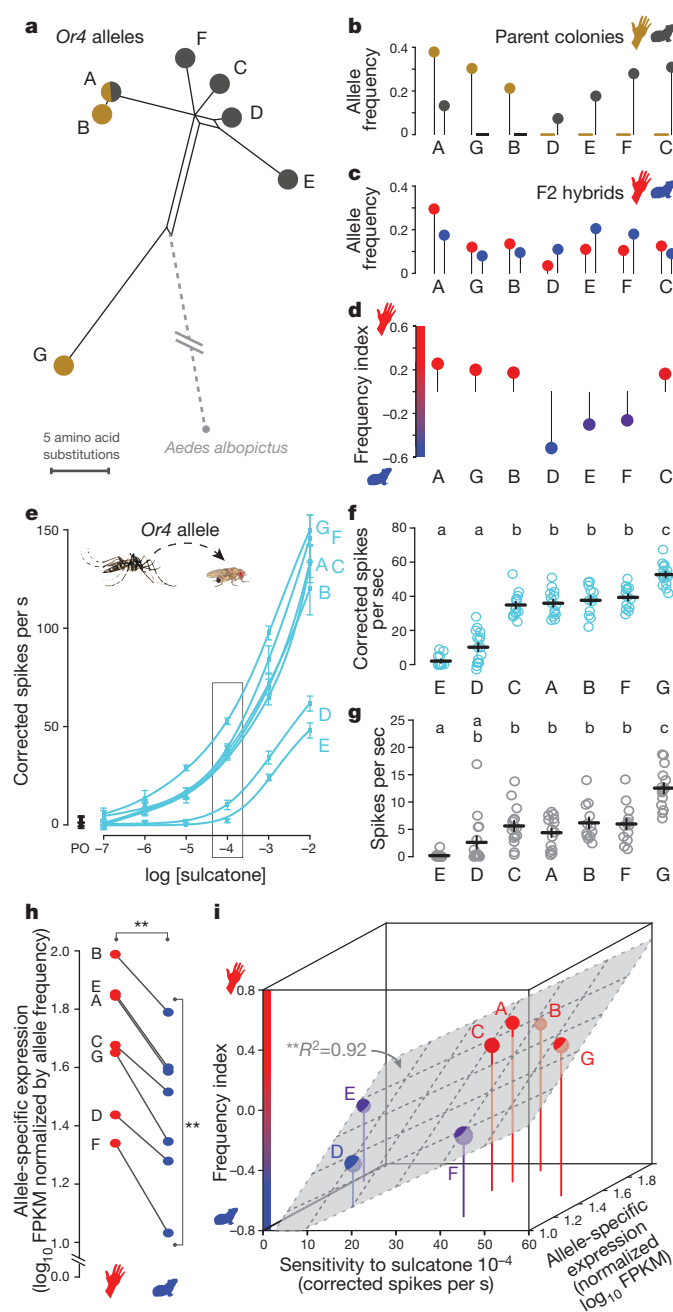


Figure 5 | Tight linkage of *Or4* allelic expression and sulcatone sensitivity to preference for humans. **a**, Protein haplotype network showing seven *Or4* alleles segregating within the K27 forest (black) and K14 domestic (brown) parent colonies with a related *Ae. albopictus* gene shown for reference. Allele A is the reference allele tested in Fig. 4a. **b**, **c**, Frequency of *Or4* alleles in K14 and K27 parent colonies ($n = 34$) (**b**) and human-preferring and guinea-pig-preferring F2 hybrids ($n = 100$) (**c**). In **c**, *Or4* alleles were inherited at significantly different frequencies by F2s with different host preference (χ^2 ; $P = 0.0002$). **d**, Relative frequency of *Or4* alleles in human-preferring and guinea-pig-preferring F2 hybrids. Data replotted from **c**. Index equal to the difference in frequency divided by the sum of the frequencies. Index bounds of -1 and 1 indicate the allele was only present in guinea-pig or human-preferring F2s, respectively. **e**, Sulcatone dose-response curves, with mean \pm s.e.m. at each concentration. PO, paraffin oil. **f**, **g**, Summary of responses to 10^{-4} sulcatone (**f**) and spontaneous activity (**g**). Alleles marked with different lowercase letters are significantly different (one-way ANOVA $P < 0.0001$ for both variables followed by Tukey's test; $n = 13$ –17 sensilla per allele). **h**, Allele-specific expression derived from human-preferring and guinea-pig-preferring F2 hybrids. Variation was significant both between F2 groups (two-way ANOVA; $P = 0.004$), and among alleles (two-way ANOVA; $P = 0.005$). **i**, Joint effects of allele-specific ligand-sensitivity and expression on the F2 frequency index. Multiple regression: $P = 0.006$ for ligand-sensitivity, $P = 0.01$ for expression, $R^2 = 0.92$.

most extreme case (Fig. 5a; Extended Data Fig. 3). Given this high level of variation, we conducted a molecular analysis to confirm that all alleles correspond to a single copy gene (Extended Data Fig. 4). We tested the function of each allele in *Drosophila* and found that A, B, C, F, and G were highly sensitive to sulcatone, whereas D and E were much less sensitive (Fig. 5e, f). Variation in spontaneous activity mirrored the variation in odour-evoked activity (Fig. 5g).

We also asked whether *Or4* alleles vary in expression and could thus help explain the upregulation of this locus in human-preferring mosquitoes. We reanalysed the RNA-seq data, parsing gene expression according to major alleles in each F2 pool and normalizing by allele frequency to isolate levels of allele-specific expression. Two major conclusions emerged. First, all alleles are expressed at higher levels when carried by human-preferring F2s than when carried by guinea-pig-preferring F2s (Fig. 5h, compare red to blue). This difference suggests that a genetic element unlinked from *Or4* contributes to upregulation in human-preferring mosquitoes. Second, the rank order of gene expression of each allele was preserved, regardless of whether it was carried by human-preferring or guinea-pig-preferring mosquitoes (Fig. 5h, compare alleles within red or blue). For example, B is always expressed at the highest level and F at the lowest level. These consistent and significant differences among alleles suggest that genetic elements linked to *Or4* and varying among alleles also contribute to changes in expression.

To determine which characteristics of *Or4* alleles are tied to behaviour, we asked whether ligand-sensitivity and/or expression can account for host preference-based differential inheritance of alleles (Fig. 5d). Remarkably, both factors had significant effects on relative allele frequency in F2s and together explained 92% of the variation (Fig. 5i). This striking relationship suggests that expression and sensitivity to sulcatone have independent and additive effects on preference. Moreover, strong human preference appears to require alleles with both high sensitivity and expression. For example, E is one of the most highly expressed alleles, yet it has weak sensitivity to sulcatone and is biased towards guinea-pig-preferring mosquitoes (Fig. 5i). Conversely, allele F has high sensitivity to sulcatone, yet it is expressed at extremely low levels and is also slightly biased towards guinea-pig-preferring mosquitoes (Fig. 5i).

Discussion

We have re-established the Rabai forest and domestic mosquito study system for investigation of the striking evolutionary adaptations that help domestic *Ae. aegypti* females find, bite, and thereby spread disease to humans. We show that preference for humans is tightly linked to increases in both the expression and ligand-sensitivity of odorant receptor *Or4*. These changes may help mosquitoes distinguish humans from non-human animals by increasing behavioural sensitivity to the signature human odorant sulcatone. Interestingly, sulcatone has been described as a mosquito repellent when added to human odour at certain concentrations^{43–45} and sometimes as an attractant when added at low concentrations or delivered alone^{29,45}. This raises the intriguing possibility that while a baseline level of sulcatone signals humanness, mosquitoes may prefer humans that have lower levels of sulcatone over those with high levels. In other words, as is true for many odours, it is possible to have too much of a good thing. We further note that sulcatone is unlikely to be the only odorant that makes us smell human, nor *Or4* the only gene contributing to human preference in domestic *Ae. aegypti*. Guinea-pig odour perfumed with sulcatone was not preferred over the odour of guinea-pig alone by human-preferring mosquitoes (Extended Data Fig. 5). We strongly suspect that evolutionary changes at other loci, including some of the other candidates from our antennal RNA-seq analysis, may also play a role.

Our results also provide insight into the molecular basis of behavioural evolution. Despite exciting progress in this area^{46,47}, examples of specific genes associated with behavioural change are extremely rare⁴⁸. Previous authors documented changes in the peripheral olfactory system of organisms with novel host preference^{16,18,19}, but direct links between these changes and behaviour have been missing. We have established

a clear genetic association between such changes and behaviour. This work begins to unravel the molecular genetic basis of an important evolutionary shift in insect host preference. More generally, such host shifts not only impact the efficiency of mosquitoes as vectors of infectious disease, but contribute to the economic damage caused by agricultural pests⁴⁹ and play a key role in the formation of new species⁵⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 January; accepted 13 October 2014.

- Lehane, M. J. *The Biology of Blood-sucking in Insects* (Cambridge Univ. Press, 2005).
- Christophers, S. R. *Aedes aegypti*, the Yellow Fever Mosquito: Its Life History, Bionomics and Structure Ch. 2 (Cambridge Univ. Press, 1960).
- Mattingly, P. F. Genetical aspects of the *Aedes aegypti* problem. I: taxonomy and bionomics. *Ann. Trop. Med. Parasitol.* **51**, 392–408 (1957).
- Brown, J. E. et al. Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. *Evolution* **68**, 514–525 (2014).
- Lumsden, W. H. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952–53. II. general description and epidemiology. *Trans. R. Soc. Trop. Med. Hyg.* **49**, 33–57 (1955).
- Trpis, M. & Hausermann, W. Demonstration of differential domesticity of *Aedes aegypti* (L) (Diptera, Culicidae) in Africa by mark-release-recapture. *Bull. Entomol. Res.* **65**, 199–208 (1975).
- Petersen, J. L. *Behavioral Differences in Two Subspecies of Aedes aegypti* (L) (Diptera: Culicidae) in East Africa. PhD thesis, Univ. Notre Dame (1977).
- Trpis, M. & Hausermann, W. Genetics of house-entering behavior in East African populations of *Aedes aegypti* (L) (Diptera: Culicidae) and its relevance to speciation. *Bull. Entomol. Res.* **68**, 521–532 (1978).
- Saul, S. H., Novak, R. J. & Ross, Q. E. The role of the preadult stages in the ecological separation of 2 subspecies of *Aedes aegypti*. *Am. Midl. Nat.* **104**, 118–134 (1980).
- Gouck, H. K. Host preferences of various strains of *Aedes aegypti* and *Aedes simpsoni* as determined by an olfactometer. *Bull. World Health Organ.* **47**, 680–683 (1972).
- McKenna, R. J. *Attraction of Seven Strains of Aedes aegypti to Man and Guinea-pig in the Laboratory*. PhD thesis, Univ. California, Davis (1973).
- Bhatt, S. et al. The global distribution and burden of dengue. *Nature* **496**, 504–507 (2013).
- Brown, J. E. et al. Worldwide patterns of genetic differentiation imply multiple ‘domestications’ of *Aedes aegypti*, a major vector of human diseases. *Proc. R. Soc. B* **278**, 2446–2454 (2011).
- Mukwaya, L. G. Host preference in *Aedes* (*Stegomyia*) mosquitoes in Uganda. II. Studies on indoor and outdoor biting and resting behaviour with special reference to *Aedes aegypti* L. *Acta Trop.* **31**, 165–176 (1974).
- Dekker, T., Geier, M. & Carde, R. Carbon dioxide instantly sensitizes female yellow fever mosquitoes to human skin odours. *J. Exp. Biol.* **208**, 2963–2972 (2005).
- Dekker, T., Ibba, I., Siju, K. P., Stensmyr, M. C. & Hansson, B. S. Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila melanogaster* sibling, *D. sechellia*. *Curr. Biol.* **16**, 101–109 (2006).
- Matsuo, T., Sugaya, S., Yasukawa, J., Aigaki, T. & Fuyama, Y. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biol.* **5**, e118 (2007).
- Linz, J. et al. Host plant-driven sensory specialization in *Drosophila erecta*. *Proc. R. Soc. B* **280**, 20130626 (2013).
- Rinker, D. C., Zhou, X., Pitts, R. J., Rokas, A. & Zwiebel, L. J. Antennal transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in *Anopheles gambiae*. *BMC Genomics* **14**, 749 (2013).
- Wada-Katsumata, A., Silverman, J. & Schal, C. Changes in taste neurons support the emergence of an adaptive behavior in cockroaches. *Science* **340**, 972–975 (2013).
- Michelsmore, R. W., Paran, I. & Kesseli, R. V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl Acad. Sci. USA* **88**, 9828–9832 (1991).
- Robertson, H. M., Warr, C. G. & Carlson, J. R. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **100** (Suppl 2), 14537–14542 (2003).
- Croset, V. et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* **6**, e1001064 (2010).
- Fan, J., Francis, F., Liu, Y., Chen, J. L. & Cheng, D. F. An overview of odorant-binding protein functions in insect peripheral olfactory reception. *Genet. Mol. Res.* **10**, 3056–3069 (2011).
- DeGennaro, M. et al. *orco* mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. *Nature* **498**, 487–491 (2013).
- Dobritsa, A. A., van der Goes van Naters, W., Warr, C. G., Steinbrecht, R. A. & Carlson, J. R. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37**, 827–841 (2003).
- Cork, A. & Park, K. C. Identification of electrophysiologically-active compounds for the malaria mosquito, *Anopheles gambiae*, in human sweat extracts. *Med. Vet. Entomol.* **10**, 269–276 (1996).

28. Labows, J., Preti, G., Hoelzle, E., Leyden, J. & Kligman, A. Analysis of human axillary volatiles: compounds of exogenous origin. *J. Chromatogr.* **163**, 294–299 (1979).
29. Bernier, U. R., Kline, D. L., Barnard, D. R., Schreck, C. E. & Yost, R. A. Analysis of human skin emanations by gas chromatography/mass spectrometry. 2. Identification of volatile compounds that are candidate attractants for the yellow fever mosquito (*Aedes aegypti*). *Anal. Chem.* **72**, 747–756 (2000).
30. Syed, Z. & Leal, W. S. Acute olfactory response of *Culex* mosquitoes to a human- and bird-derived attractant. *Proc. Natl Acad. Sci. USA* **106**, 18803–18808 (2009).
31. Birkett, M. A. *et al.* The role of volatile semiochemicals in mediating host location and selection by nuisance and disease-transmitting cattle flies. *Med. Vet. Entomol.* **18**, 313–322 (2004).
32. Jumean, Z., Gries, R., Unruh, T., Rowland, E. & Gries, G. Identification of the larval aggregation pheromone of codling moth, *Cydia pomonella*. *J. Chem. Ecol.* **31**, 911–924 (2005).
33. Nielsen, B. L. L., Jérôme, N., Saint-Albin, A., Rampin, O. & Maurin, Y. Behavioural response of sexually naive and experienced male rats to the smell of 6-methyl-5-hepten-2-one and female rat faeces. *Physiol. Behav.* **120**, 150–155 (2013).
34. Stoeffler, M., Maier, T. S., Tolasch, T. & Steidle, J. L. M. Foreign-language skills in rove-beetles? Evidence for chemical mimicry of ant alarm pheromones in myrmecophilous *Pella* beetles (Coleoptera: Staphylinidae). *J. Chem. Ecol.* **33**, 1382–1392 (2007).
35. Rines, H. W., French, R. C. & Daasch, L. W. Nonanal and 6-methyl-5-hepten-2-one: endogenous germination stimulators of uredospores of *Puccinia graminis* var. *tritici* and other rusts. *J. Agric. Food Chem.* **22**, 96–100 (1974).
36. Socaci, S. A. *et al.* Chemometric discrimination of different tomato cultivars based on their volatile fingerprint in relation to lycopene and total phenolics content. *Phytochem. Anal.* **25**, 161–169 (2013).
37. Webster, B. *et al.* Identification of volatile compounds used in host location by the black bean aphid, *Aphis fabae*. *J. Chem. Ecol.* **34**, 1153–1161 (2008).
38. Carey, A. F., Wang, G., Su, C. Y., Zwiebel, L. J. & Carlson, J. R. Odorant reception in the malaria mosquito *Anopheles gambiae*. *Nature* **464**, 66–71 (2010).
39. Bohbot, J. *et al.* Molecular characterization of the *Aedes aegypti* odorant receptor gene family. *Insect Mol. Biol.* **16**, 525–537 (2007).
40. Krzywinski, J., Grushko, O. G. & Besansky, N. J. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol. Phylogenet. Evol.* **39**, 417–423 (2006).
41. Pellegrino, M., Steinbach, N., Stensmyr, M. C., Hansson, B. S. & Vosshall, L. B. A natural polymorphism alters odour and DEET sensitivity in an insect odorant receptor. *Nature* **478**, 511–514 (2011).
42. Keller, A., Zhuang, H., Chi, Q., Vosshall, L. B. & Matsunami, H. Genetic variation in a human odorant receptor alters odour perception. *Nature* **449**, 468–472 (2007).
43. Logan, J. G. *et al.* Arm-in-cage testing of natural human-derived mosquito repellents. *Malar. J.* **9**, 239 (2010).
44. Logan, J. G. *et al.* Identification of human-derived volatile chemicals that interfere with attraction of *Aedes aegypti* mosquitoes. *J. Chem. Ecol.* **34**, 308–322 (2008).
45. Menger, D. J., Van Loon, J. J. A. & Takken, W. Assessing the efficacy of candidate mosquito repellents against the background of an attractive source that mimics a human host. *Med. Vet. Entomol.* <http://dx.doi.org/10.1111/mve.12061> (2014).
46. Greenwood, A. K., Wark, A. R., Yoshida, K. & Peichel, C. L. Genetic and neural modularity underlie the evolution of schooling behavior in threespine sticklebacks. *Curr. Biol.* **23**, 1884–1888 (2013).
47. Weber, J. N., Peterson, B. K. & Hoekstra, H. E. Discrete genetic modules are responsible for complex burrow evolution in *Peromyscus* mice. *Nature* **493**, 402–405 (2013).
48. Martin, A. & Orgogozo, V. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**, 1235–1250 (2013).
49. van der Goes van Naters, W. & Carlson, J. R. Insects as chemosensors of humans and crops. *Nature* **444**, 302–307 (2006).
50. Coyne, J. A. & Orr, H. A. *Speciation* Ch. 4 and 5 (Sinauer Associates, 2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. K. N. Lawniczak, K. J. Lee, M. N. Nitabach, and the Vosshall laboratory for discussion and comments on the manuscript; J. E. Brown and J. R. Powell for discussion and coordination of field collections; W. Takken for advice regarding many aspects of this work; J.-P. Mutebi, B. Miller, and A. Ponlawat for live specimens from Uganda and Thailand; D. Beck, K. Nygaard, K. Prakash, and L. Seeholzer for expert technical assistance. We also thank X. Chen for pre-publication access to a draft *Ae. albopictus* genome assembly, and J. Liesch for access to Orlando strain RNA-seq data. We received valuable advice on collecting and working with forest and domestic forms of *Ae. aegypti* from M. Trpis, J. L. Peterson, and P. Lounibos, and on the design and use of two-port olfactometers from U. Bernier and V. Sherman. This work was funded in part by a grant to R. Axel and L.B.V. from the Foundation for the National Institutes of Health through the Grand Challenges in Global Health Initiative. This work was supported in part by the following National Institutes of Health grants: K99 award from NIDCD to C.S.M. (DC012069), an NIAID Vectorbase DBP subcontract to L.B.V. (HHSN272200900039C), and a CTSA award from NCATS (5UL1TR000043). R.I. received support from the Swedish Research Council and SLU: Insect Chemical Ecology and Evolution (IC-E3). L.B.V. is an investigator of the Howard Hughes Medical Institute.

Author Contributions C.S.M. and L.B.V. conceived the study. C.S.M. participated in the execution and analysis of all aspects of the study. J.L. helped coordinate mosquito collection in Rabai, Kenya under the supervision of R.S. S.A.S. helped design and carry out the morphological assays presented in Fig. 1e–i. F.B. helped clone, analyse, and genotype mosquitoes for the *Or4* alleles presented in Fig. 5a–d, and construct transgenic *Drosophila* lines for use in single sensillum recordings. A.B.O. and R.I. designed, conducted, and analysed the GC–SSR and GC–MS experiments presented in Fig. 4 and carried out pilot experiments comprising dose–response curves and spontaneous activity analysis of alleles A and E, similar to those presented in Fig. 5e–g. C.S.M. and L.B.V. designed all other experiments, interpreted the results, designed the figures, and wrote the paper.

Author Information Raw RNA-seq data are available for download at the NCBI Sequence Read Archive (accession number SRP035216). Coding sequences of *AegOr4* alleles are at GenBank (accession numbers KF801614, KF801615 and KF801617–KF801621). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.B.V. (Leslie.Vosshall@rockefeller.edu).

METHODS

Ethics and regulatory information. Mosquitoes were collected and exported from Kenya with approval of the director of the Kenya Medical Research Institute under the study approved by the Scientific Steering Committee and Ethical Review Committee (SSC No. 1679). Live mosquito eggs were imported to the USA with permits issued by the United States Department of Agriculture and the Centers for Disease Control and Prevention. The use of non-human animals in host preference tests at The Rockefeller University was approved and monitored by The Rockefeller University Institutional Animal Care and Use Committee (Protocol 11487). The participation of humans in blood-feeding mosquitoes during routine colony maintenance and as subjects in host preference tests at The Rockefeller University was approved and monitored by The Rockefeller University Institutional Review Board (IRB protocol LVO-0652). The protection of human subjects and ethical work with animals at the Swedish University of Agricultural Sciences (SLU) was in accordance with the Central Ethical Review Board and the Committee for Laboratory Animal Science in Sweden, respectively. All human subjects gave their informed consent to participate in the work carried out at The Rockefeller University and at SLU.

Field collection and creation of laboratory colonies. We collected mosquito (*Ae. aegypti*) eggs, larvae, and pupae in indoor and outdoor habitats in the Rabai region of Kenya in January 2009 (39° 34'–36° E, 3° 55'–57' S). Indoor collections in the Rabai region of Kenya (Fig. 1a) were made, with the verbal permission of homeowners, by visually scanning artificial containers used to store water with flashlights and removing larvae and pupae with a nylon sieve. Artificial containers harbouring mosquitoes included plastic buckets, metal jerry cans, and traditional earthenware pots (Fig. 1b). Outdoor collections were made in both village environments and nearby forest fragments along the Kombeni River (Fig. 1a) in two ways. First, larvae and pupae were removed from artificial containers left outdoors using a sieve, and from natural containers such as tree holes (Fig. 1c, top) using a turkey baster or small plastic pipette. In some cases, tree holes retained water from the last rains, while in others we introduced well water two days before collection to induce hatching of dormant eggs. Second, freshly laid eggs were collected in oviposition traps nailed to trees (Fig. 1c, bottom) or left on the ground for 3 days. Traps comprised black plastic cups (13 cm diameter, 15 cm tall) half-filled with water and lined with coarse brown seed-germination paper (76 pound, Anchor paper) (Fig. 1c, bottom). Adult females attracted to the cups laid their eggs on the wet paper, which was then removed and dried to prevent embryos from hatching. When mosquitoes were collected as larvae or pupae, individuals originating in the same home/container or cluster of nearby homes/containers were reared to adulthood, mated with each other, blood-fed, and induced to oviposit in a field laboratory so that eggs of the first laboratory generation could be dried and exported to the USA. A total of 29 laboratory colonies were established (K1–K29), each founded by between 1 and 14 females (median 4) collected within 0–50 m of each other in the field (Fig. 1a) and mated to males collected from the same area. The only exception was K27, which was established using males from a forest tree hole that yielded no females. We therefore mated the males with females collected outdoors in a village 2 km away (Fig. 1a). The number of females that had the opportunity to contribute to the six Kenyan colonies characterized in Figs 1g–i, 2g, h, and 3c, d is as follows: K2 ($n = 3$), K4 ($n = 6$), K14 ($n = 2$), K18 ($n = 2$), K19 ($n = 6$), K27 ($n = 9$). However, some of these females may not have laid eggs and thus ultimately not contributed genetic material to the colony. We also established one colony with eggs sent to us from Bundibugyo, Uganda and another with eggs sent from Rayong, Thailand. We denote the generation of a laboratory colony with lower case 'g' followed by a number; for example, g1 refers to the first laboratory generation.

Colony maintenance and insect rearing. Mosquitoes were maintained at 25–28 °C with 70–80% relative humidity under a 14 h light: 10 h dark cycle (lights on at 8 a.m.). Eggs were hatched by submersion in a broth containing deoxygenated deionised water and powdered Tetramin tropical fish food (Tetra). Larvae were cultured in deionised water and fed Tetramin tablets. Adults were maintained in large plastic cages (30 × 30 × 30 cm, BioQuip) and given unlimited access to 10% sucrose. In each generation, 75–250 adult females from each colony were blood-fed on a human volunteer. To minimise the potential for natural selection on host preference, a human arm was offered 1–2 times each day until 90% or more of the females had taken a full blood-meal. Eggs were collected in 96 ml black plastic soufflé cups (Solo cup company) lined with seed-germination paper and filled with 30 ml of 'soil water' prepared by incubating deionised water with commercial potting soil at room temperature in an open vessel for 1–10 weeks. Eggs were dried to prevent hatching and stored at 18 °C, 85% relative humidity for 6–12 months. Eggs from domestic colonies K1–K6 were less tolerant to drying and storage than eggs from forest colonies. They were therefore hatched after only 1–4 months, resulting in shorter generation times.

Morphological analysis. We characterized the scale and cuticle colour of mosquitoes from 14 colonies in 2010 (g3 to g7; $n = 10$ –15 mosquitoes per colony), and a partially overlapping set of 8 colonies in 2012 (g5 to g11; $n = 10$ –15 mosquitoes per

colony). We collected adult female mosquitoes within 48 h of eclosion and stored them at –20 °C for up to 4 weeks. Immediately upon removal from the freezer, we inserted an insect pin laterally through the thorax and positioned each specimen using a pinned specimen manipulator (Rose Entomology) one at a time under a microscope (Nikon Eclipse SMZ1500) fitted with a ring light and a digital camera (Nikon digital sight DS-2Mv controlled by NIS-Elements F v3.0 software). We took photographs in two characteristic positions, one highlighting the dark and light scales that decorate the scutum (Extended Data Fig. 1a) and the other showing the bare cuticle of the postnotum on the posterior face of the thorax plus the first 3–4 segments of the abdomen (Extended Data Fig. 1b). We measured scale and cuticle colour on the dark parts of the scutum and postnotum, respectively, in Adobe Photoshop (v. CS6, Adobe Systems Inc.) by sampling RGB values at four characteristic positions using the colour picker tool (Extended Data Fig. 1a, b), and converted RGB numbers to hue, saturation, value (HSV) using R software (v. 2.15.0, <http://www.r-project.org/>). The light environment and camera settings were held constant for the duration of each analysis, but differed slightly between analyses, such that values from 2010 and 2012 are not directly comparable.

We assessed the extent of white scaling on the first abdominal tergite in the 2012 analysis using an ordinal scale from 1 to 5 (Extended Data Fig. 1c) as follows: 1, up to a few scattered white scales; 2, small patch of white scales at midline; 3, contiguous patch of white scales at midline stretching from top to bottom of tergite and covering up to 60% of visible area; 4, contiguous patch of white scales covering 60–90% of visible area; 5, contiguous patch of white scales covering > 90% of visible area. These scores correspond roughly to letters F through K in the scheme of McClelland^{45,52}. A single experimenter scored all mosquitoes blind to their identity.

Host preference assays. Mosquitoes used in host preference assays were adult females 1–3 weeks post eclosion that had been given the opportunity to mate, but had not taken a blood-meal. Females were sorted briefly under cold anaesthesia (4 °C) and deprived of access to food or water 16–24 h before testing.

Biting assay. This assay was used to test mosquito preference for human versus guinea-pig (Fig. 2a) and was based on a previously described landing assay²³. For each trial, we allowed approximately 50 females to acclimate overnight in a large custom-made cage (61 × 61 × 91.5 cm) constructed with aluminium screening on 3 sides and clear vinyl on the fourth side for easy viewing (BioQuip). The following morning, we simultaneously introduced a human arm (33-year-old female) and an anaesthetized guinea-pig (pigmented strain, one of two females, 2–6 months old) through cloth sleeves at opposite ends of the cage and rested them on the floor of the cage ~60 cm apart (Fig. 2a), and recorded the number of mosquitoes that blood-fed within 10 min. We defined blood-feeding as landing on the host, inserting the proboscis, and drawing enough blood into the abdomen that it was visible to the naked eye of the observer. Preference index was calculated as number of mosquitoes that blood-fed on the human minus the number that blood-fed on the guinea-pig divided by the total number of mosquitoes that blood-fed on either host. Overall response was the fraction of mosquitoes that blood-fed on either host. We tested g1–g2 females from 16 colonies, 1–3 trials per colony and assessed the significance of the difference between forest and domestic colonies using two-sided, two-sample *t*-tests, where each colony served as a single data point.

Live host olfactometer assay. We used a live host olfactometer assay to test mosquito preference for human versus non-human animals in the absence of visual cues, as previously described²⁵ (Fig. 2d). A total of 50–100 females acclimated in a dual-port olfactometer for 15 min. We then activated a fan and opened a sliding door, exposing the mosquitoes to streams of air that had passed over the arm of a human volunteer (one of three females, 22–35 years old) or an awake guinea-pig (pigmented strain, one of two females, 6–24 months old) or chicken (domestic strain, sex unknown, 2–4 weeks old). During an 8 min trial, mosquitoes choosing to fly upwind towards the odour of either host were trapped in small ports. The breath of both hosts provided a source of carbon dioxide (CO₂) in their respective air streams. The comfort of the animal was assured by rest and feeding periods outside of the olfactometer between trials and the minimization of any stress while in the olfactometer. Neither the animals nor the human volunteer were bitten by mosquitoes during this assay. We screened g1–g5 females from 26 colonies, 1–7 trials per colony in the human versus guinea-pig comparison and a subset of 5 colonies, 1 trial per colony for human versus chicken. A preference index equal to the number of mosquitoes entering the human trap minus those that entered the animal trap divided by the total number of mosquitoes entering either trap was calculated. Overall response was the fraction of mosquitoes that entered either trap. We assessed the significance of the difference between forest and domestic colonies using two-sided, two-sample *t*-tests, where the mean for each colony served as a single data point.

Nylon sleeve olfactometer assay. We substituted host-scented nylon sleeves for live hosts in the olfactometer to isolate the effects of specific host odours on mosquito preference in the context of a controlled amount of CO₂ (Fig. 2h, top). This assay was conducted as described²⁵, except that instead of pushing carbon-filtered air through the olfactometer with pumps, ambient air was gently pulled over the nylon

sleeves and into the main compartment with a fan in exactly the same way as the live host assay. Both air streams were supplemented with an equal amount of CO₂, resulting in a final concentration of 0.2–0.3%. Human- and guinea-pig-scented sleeves comprised 35-cm long sections of women's sheer nylon stockings that had been worn on a human arm (one of two females 22–35 years old) or guinea-pig torso (pigmented strain, one of two females, 6–24 months old) for 24 h during normal daily activity. Conditioned sleeves were stored at –20 °C for up to a month before use, and individual sleeves were used in only one trial per colony. We tested g2–g8 females from 6 colonies, 4–17 trials per colony. We quantified preference and overall response as described for the live host olfactometer assay. We used a one-way ANOVA followed by Tukey's test to assess significance of variation among colonies.

Sulcatone perfuming experiments. The olfactometer apparatus was modified from the main figures as follows: the air stream was carbon-filtered and simultaneously pushed by a pump and pulled by an exhaust fan. The stream first mixed with CO₂ and picked up the odour of a guinea-pig-worn nylon sleeve in a large cylinder before entering a plastic bag and splitting into two streams. The two streams then passed through two smaller cylinders containing a 55 mm diameter filter circle (Whatman) loaded with 50 µl of either solvent (paraffin oil) or 10⁻⁴ sulcatone, a concentration chosen because this is within the range of concentrations eluted from human-worn nylon sleeves and above the response threshold of most *AeegOr4* alleles as assessed in the *Drosophila* ab3A neuron. The air then entered two mosquito traps and flowed into a large mosquito chamber. In each trial, 50 1–4-week-old female mosquitoes from the human-preferring Orlando (ORL) lab strain or human-preferring K14 Rabai strain were given a choice between the guinea-pig odour/CO₂ mix supplemented with solvent on one side and sulcatone (10⁻⁴) on the other side. We used a one-sample *t*-test to evaluate if preference was different from zero or no preference. In the test, -1 indicates strong preference for the solvent side and +1 indicates strong preference for sulcatone.

Isolation of human- and guinea-pig-preferring F2 hybrids. We generated a large population of F2 mosquitoes by crossing ~150 individuals from domestic colony K14 (g5) with ~150 individuals from forest colony K27 (g5) in both directions and then interbreeding ~2,000 of their F1 progeny (Fig. 3a). We chose colonies K14 and K27 as parents because they showed the most extreme preferences in the nylon sleeve assay (Fig. 2h). We reared ~2,500 F2 females and identified the most human-preferring and guinea-pig-preferring among them using repeated tests in the live host olfactometer. F2 individuals derived from the reciprocal parent crosses were pooled for testing. The live host olfactometer was used as described above, except that human arm odour was not supplemented with human breath, and 200–300 females were tested at a time. In the first round of testing, mosquitoes that entered the human or guinea-pig trap were separated. In subsequent rounds, mosquitoes that had previously responded to human were tested separately from mosquitoes that previously responded to guinea-pig, and only individuals that responded to the same host as in previous trials were retained. Mosquitoes were allowed to rest in the insectary with unlimited access to 10% sucrose for 48 h between trials. This process resulted in a pool of 141 females that responded to the human three times in a row, and 117 females that responded to the guinea-pig twice in a row. We then returned the pools to the olfactometer one last time to quantify their preference (Fig. 3b), but retained all females in each pool regardless of their behaviour in this final test. We also tested individuals from the two parent colonies each day as positive controls. Each F2 hybrid pool was split into two equal-sized groups for the preparation of replicate RNA-seq libraries.

RNA-seq experiments. Antennal RNaseq libraries were prepared for the two F2 hybrid pools and 8 colonies using Illumina TruSeq RNA Sample Prep Kit v.2 and mRNA-Sequencing Sample Prep Kit, respectively. Libraries were sequenced on Illumina GEX and HiSeq 2000 sequencing systems. We prepared antennal RNaseq libraries for the two F2 hybrid pools and the following colonies: K2 (g8), K4 (g8), K14 (g5), K18 (g5), K19 (g4), K27 (g5), Thailand (g4), and Uganda (g3). Colony females were 2 weeks post eclosion, had been given the opportunity to mate, but had not taken a blood-meal. F2 hybrids had additionally experienced 3–4 rounds of preference testing and were 2–3 weeks post eclosion. One library was prepared for each colony ($n = 8$ libraries total), while two replicate libraries were prepared for each F2 hybrid pool ($n = 4$ libraries total), split for this purpose into two equal-sized groups as described above. For each library, we removed whole antennae, including pedicel and flagellum, from 55–150 cold-anaesthetized females with fine forceps under a stereo-microscope and placed them in a microcentrifuge tube held at -76°C in an ethanol/dry ice batch. We stored tubes at -80°C and then extracted total RNA using an RNeasy kit (Qiagen). For colonies, we prepared sequencing libraries from 1–3 μg total RNA with an mRNA-Sequencing Sample Prep Kit (Illumina). We selected 200 base pair (bp) inserts on a 2% agarose gel both before PCR enrichment, per kit instructions, and after PCR enrichment, to further narrow the insert size distribution. For F2 hybrid pools, we prepared sequencing libraries from 1.5 μg total RNA using a TruSeq RNA Sample Prep Kit v.2 (Illumina) with a 4 min fragmentation time, resulting in a broad range of insert sizes (mean ~ 330 bp, s.d. ~ 90 bp).

Colony libraries were each sequenced in one lane of an Illumina GEX and one lane of an Illumina HiSeq 2000, generating 20–30 million 76 bp and 65–90 million 101 bp single-end reads respectively. Barcoded F2 hybrid libraries were pooled, and the resulting pool was sequenced in four lanes of an Illumina HiSeq 2000, generating 175–240 million 101 bp single-end reads per library.

We used TopHat2 v.2.0.9 (ref. 54) to align unfiltered RNA-seq reads to the *Ae. aegypti* AeGL1.3 reference transcriptome, custom-revised to include community annotations for four chemosensory gene families: olfactory receptors³⁹, ionotropic receptors²³, gustatory receptors⁵⁵, and odorant binding proteins⁵⁶. We chose lenient alignment parameters allowing 3 mismatches per segment, 12 mismatches per read, and gaps of up to 3 bp (-N 12-segment-mismatches 3-read-gap-length 3-read-edit-dist 15) to minimise the possibility that sequence divergence between forest or domestic mosquitoes and the reference genome would bias expression estimates. These settings resulted in the successful alignment of 47–56% of raw reads (43–62 million per colony). Default alignment parameters allowing 2 mismatches per read and gaps up to 2 bp resulted in the same qualitative conclusions. We then used Cuffdiff2 v.2.1.1 (ref. 57) to test for differential expression between forest and domestic colonies ($n = 3$ –4 colonies each with each colony treated as a biological replicate), and between human-preferring and guinea-pig-preferring F2 hybrids ($n = 2$ replicate libraries each) with the multiple mapping correction and a false discovery rate set at 0.05. For colonies, we compared all four forest colonies (K18, K19, K27, Uganda) to all four domestic colonies (K2 K4, K14, Thailand) or just the three most guinea-pig-preferring forest colonies (K19, K27, Uganda) to the three most human-preferring domestic colonies (K4, K14, Thailand). Results were similar and we present only the latter comparison in Fig. 3c. However, data from the two excluded colonies (K2 and K18) are used in Extended Data Fig. 2. We explored and visualized data with CummeRbund v.2.0.0 (<http://compbio.mit.edu/cummeRbund/>) and custom R scripts (R software v.2.15.0, <http://www.r-project.org/>).

Transgenic *Drosophila* strains. *UAS-AeegOr4* contained the full length coding sequence of each major allele cloned into the EcoRI site of the *pUAST attB* vector⁵⁸ using the In-Fusion HD Cloning Kit (Clontech). Plasmids are available from Addgene.org as plasmid numbers 59792–59798. Constructs were injected into the *Drosophila melanogaster* ZH-86Fb strain [*P{ry+|t.7.2} = hsp70-*flp*1, y[1] w[*]; M{3xP3-RFP.*attP*}/ZH-86Fb; M{*vas-int.B*}/ZH-102F] (ref. 58) using the phiC31-based integration system targeting the 86Fb-attP docking site on chromosome III by Genetic Services Inc. *UAS-AeegOr4* transgenic fly strains are available from the Bloomington *Drosophila* Stock Center (NIH P40OD018537) as stock numbers 56143–56149. *UAS-AeegOr4* transgenes were crossed into a strain lacking *DmelOr22a/b* (*Dhalo*) and carrying *Or22a-GAL4* (refs 26, 59). Electrophysiological recordings were obtained from flies of genotype *w;Dhalo/Dhalo;Or22a-GAL4/UAS-AeegOr4*.*

Human and animal volatile collections. We collected headspace volatile extracts from nylon sleeves previously worn by each of 5 human volunteers (3 female, 22–47 years old) and 4 guinea-pigs for 24 h (see nylon sleeve host preference assay), and unworn control sleeves (Fig. 4c, d). Four sleeves per human or guinea-pig or control replicate were enclosed in a 0.5 litre glass jar. A charcoal filtered air stream ($1.5 \text{ litre min}^{-1}$) was drawn by a diaphragm vacuum pump (KNF Neuberger) over the sleeves, from the bottom to the top of the jar, onto an air filter for 3 h. The air filters were made of Teflon tube ($4 \times 50 \text{ mm}$), holding 50 mg Porapak Q adsorbent (80/100 mesh, Altech) between glass wool plugs. The filters were rinsed with 4 ml acetone ($>99.9\%$ pure, HPLC Grade, Chromasolv Plus, Sigma-Aldrich), redistilled ethanol and pentane ($>99\%$ pure, GC grade, Sigma-Aldrich) before use. Adsorbed volatiles were desorbed by eluting with 500 μl pentane, and condensed to 20% of their original volume, under a stream of nitrogen. Heptyl acetate (50 ng, 99.8% pure, Sigma-Aldrich) was added as an internal quantification standard. Two negative technical controls were included: a filter control sample obtained by running an elution step on cleaned adsorbent filters and a solvent-only control.

We also collected headspace volatiles from live human volunteers and live non-human animals or hair from non-human animals. Human body volatiles were collected as described⁶⁰ by placing naked volunteers in customised heat-sealed cooking bags, introducing synthetic air into the bags, and extracting it with pumps through columns containing Porapak Q for a period of 2.5 h ($n = 20$ humans). Volatiles were collected from one live chicken placed on a metal mesh in an airtight desiccator covered with a black cloth. Charcoal filtered air was introduced (1 litre min^{-1}) via a Teflon tube and pumped out of the desiccator via a glass splitter connected to four Porapak Q adsorbent columns ($0.25 \text{ litre min}^{-1}$ each) for 1 h. Volatiles from cattle, horse, and sheep were collected by placing 20 g of freshly shaved hair or wool from multiple individuals in a 0.5 litre glass wash bottle ($n = 10$ cattle, $n = 2$ horses, $n = 5$ sheep). Charcoal filtered air was drawn by pumps ($0.1 \text{ litre min}^{-1}$) through the bottle onto a Porapak Q adsorbent column over 24 h. Before use, the adsorbent columns were rinsed with 1 ml each of methanol, dichloromethane, and pentane. Trapped volatiles were desorbed by eluting each column with $600 \mu\text{l}$ of pentane (puriss p.a., Sigma-Aldrich Chemie GmbH). Heptyl acetate was added to each extract as an internal quantification standard (500 ng for human, $2 \mu\text{g}$ for

chicken, 100 ng for other animals). Extracts from the 20 individual humans were pooled before the addition of the internal standard. Final collections were concentrated under a gentle stream of nitrogen before analysis.

Gas chromatography-coupled single sensillum recording (GC-SSR). We used GC-SSRs as described⁶¹ to screen human and guinea-pig odour for individual volatiles that activated *Or4* allele A, which was heterologously expressed in the *Drosophila* ab3A neuron (Fig. 4a, b). We separated volatiles from host odour extracts using an Agilent 6890 gas chromatograph (GC) (Agilent Technology) fitted with a fused silica capillary column (30 m × 0.25 mm i.d.) coated with non-polar HP-5 stationary phase (d.f. = 0.25 µm), and using hydrogen gas as the mobile phase (45 cm s⁻¹). Aliquots of the extracts (5–7.5 µl) were injected splitless for 30 s, with the injector maintained at 225 °C. The GC oven temperature was programmed from 30 °C (3 min hold), followed by a ramp of 8 °C min⁻¹ to 225 °C, and held isothermal for 10 min. The GC was fitted with a make-up gas fed (4 psi N₂) four-way cross (Graphpack 3D/2 Crosspiece Sulfinert, Gerstel) at the end of the column, delivering half of the effluent to the flame ionization detector and the other half to the air stream passing over the antenna of an immobilised fly via a Gerstel ODP-2 transfer line maintained at 135 °C for 15 min and increased at 8 °C min⁻¹.

We monitored the response to GC-separated volatiles of *Or4*-expressing ab3A neurons of a fly restrained under a Nikon Eclipse microscope (E600-FN8). Using a piezoelectric micromanipulator (DC-3K, Märzhäuser), an electrolytically sharpened tungsten microelectrode was introduced into the shaft or base of an ab3 sensillum and the reference tungsten electrode was inserted into the eye of the fly. The recording electrode was connected to a preamplifier (×10, Syntech) and the electrical signals were fed through an analogue-digital signal converter (IDAC-4, Syntech) and then visualized and recorded on a computer using Autospike software (Syntech). The mounted fly was placed in a continuous humidified charcoal-filtered airstream, into which GC-separated volatiles were introduced, delivered at 1 m s⁻¹ via a glass tube (6 mm i.d.).

We performed 3 replicates of GC-SSR for each of the following host odour collections: human body headspace, human-worn sleeve headspace, guinea-pig-worn sleeve headspace. Compounds were considered bioactive if their elution corresponded to changes in ab3A neuron activity in all 3 replicates. Bioactive compounds were then identified via gas chromatography-mass spectrometry (GC-MS) as described below.

Chemical analysis of volatile collections. We used gas chromatography-mass spectrometry (GC-MS) to identify the single bioactive component of human sleeve odour noted in GC-SSR experiments. Human sleeve volatile extract was injected (2 µl) into a combined Agilent 6890N gas chromatograph and 5975 mass spectrometer (Agilent Technology) fitted with an HP-5 column and programmed as for the GC-SSR analyses. The active compound was identified by comparison with reference mass spectra in our custom made library and commercially available libraries (NIST05 and Wiley). The putative identification of sulcatone was confirmed by parallel injections of synthetic reference compounds with authentic samples on the GC-MS. We also used GC-MS to quantify the sulcatone and benzaldehyde content of human and animal volatile collections by ratio of their corresponding peak areas to that of the internal standard, heptyl acetate. Absolute quantities were divided by the total time over which the respective collections were made to calculate emission rate. Emission rates of human, guinea-pig, and control sleeve extracts were compared using pairwise non-parametric Wilcoxon tests with Bonferroni correction.

***Or4* cDNA cloning.** We prepared two antennal cDNA libraries for each parent colony. For the first library we started with the same total antennal RNA used to prepare RN-seq libraries. We treated the RNA with DNase using the TURBO DNA-free Kit (Ambion) and synthesized single-stranded cDNA using Superscript III (Invitrogen) according to instructions. For the second library, we started with total antennal RNA prepared as described for RNA-seq experiments, except the dissected females came from the subsequent generation (K14 g6, K27 g6). We again treated with DNase, but then isolated mRNA using the Oligotex mRNA Mini Kit (Qiagen) before synthesizing cDNA with Superscript III. We amplified the full coding sequence of *Or4* from each cDNA library with AccuPrime Taq DNA polymerase (Invitrogen) using primers designed to anneal to the 5' and 3' untranslated regions: 5'-CGGAGTTTCCTTCGTCAGAA-3' (forward), 5'-TCGACCACTCCTATACATCGC-3' (reverse). Gel-extracted PCR amplicons (MinElute Gel Extraction Kit, Qiagen) were then cloned into pCR4 using the TOPO TA Cloning Kit (Invitrogen) and sequenced (Genewiz Inc.). We sequenced 48 clones for each cDNA library, or 96 clones per parent colony. All 192 sequences fell within one of 8 distinct allele classes, 7 of which were subsequently named alleles A–G (see below). 3' RACE reactions produced no additional sequences, providing further evidence that these alleles were representative of parent colony diversity.

***Or4* haplotype network and outgroup.** We used the program SplitsTree⁶² to infer the haplotype network shown in Fig. 5a using the distance-based SplitDecomposition algorithm. The outgroup sequence from *Aedes albopictus* was identified via blast search of allele A to an *Ae. albopictus* draft genome assembly, which was partly

supported by the Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2009) (Xiaoguang Chen, personal communication).

***Or4* genotyping and definition of major *Or4* alleles.** Single Molecule Real Time (SMRT) sequencing (Pacific Biosciences) of barcoded PCR amplicons was used to genotype 34 females from each parent colony and 100–102 females from each F2 hybrid pool. DNA was extracted from bodies (parent colonies) or single legs (F2 hybrids) using a DNeasy Blood & Tissue Kit (Qiagen). The individuals providing DNA were a subset of those whose antennae had been dissected for RNA-seq. We then amplified a 786 bp fragment of *Or4* with KOD Hot Start DNA Polymerase (Novagen) using barcoded primers complementary to invariant regions of the second and third coding exons: 5'-barcode-GTTGACCTATTGCGTTTCG-3' (forward), 5'-barcode-GCACATCAGAACAGAACTTGC-3' (reverse). The 48 forward and 48 reverse barcode sequences, provided by Pacific Biosciences, were 16 bp long and custom-paired in 92 unique combinations. We were thus able to create 3 independent pools, each containing the purified PCR amplicons of up to 92 mosquitoes. A sequencing library was prepared from each pool and sequenced on 1–2 SMRT cells of a Pacific Biosciences RS Sequencer by staff at the Cold Spring Harbour Laboratory's PacBio core facility. We obtained 20,000–30,000 of 750–850 bp circular consensus sequence (CCS) reads per pool.

We deconvoluted CCS read barcodes using the PacBioBarcodeIDCCS.py python script (<https://github.com/PacificBiosciences/DevNet/wiki/Find-Barcodes-in-CCS-Reads>), retaining only those reads with high quality assignments (barScore > 38). This resulted in an average of 158 reads per individual. We then assigned reads to alleles via blastn search⁶³ to a custom database containing the eight *Or4* coding sequences cloned from cDNA. As SMRT sequencing has a relatively high error rate, we did not require a perfect match between reads and assigned alleles, instead simply assigning reads to the allele with the best blastn score. Indeed, a fraction of reads from all mosquitoes had mismatches with respect to their assigned alleles, but the median number of mismatches per individual was almost always zero. In a few cases, individuals had reads that differed from the alleles to which they had been assigned by a consistent number of mismatches, suggesting real variation rather than sequencing error. We determined that these corresponded to rare alleles not identified by cDNA cloning and represented at a frequency of < 0.1 in both parents and F2 pools and did not study them further.

The genotyping strategy described above could not differentiate two closely related alleles, subsequently named A and B (see below) that differed by a single SNP located outside the PCR amplicon. We therefore conducted secondary genotyping on all parent and F2 mosquitoes that were shown to carry one of these two alleles (*n* = 152). For each individual, we used a LightScanner (Idaho Technology Inc.) to characterize the melting curves of a 196 bp PCR amplicon from a 3'-blocked, unlabelled, 21 bp oligonucleotide probe complementary to allele B in the region surrounding the diagnostic SNP. Melting curves were characteristic to genotype, allowing easy discrimination of individuals carrying 0, 1, or 2 copies of allele B. Reactions were prepared and analysed according to manufacturer instructions using the following primers and probe: 5'-CAAGGTCTTGCAAATGATCGGTAA-3' (forward), 5'-CGATGTTGATGATCTGACCGAAA-3' (reverse), 5'-AAGTCCAGTTCGGTTTCGTG amino-modifier-3' (probe).

Seven of the 8 full-length *Or4* alleles cloned from cDNA were present at a frequency of ≥ 0.1 in at least one parent colony or F2 hybrid pool. We defined these as major alleles and named them A through G in order of increasing distance from the genome reference allele A, and deposited them in GenBank (accession numbers KF801614–KF801615 and KF801617–KF801621). The eighth allele discovered by cDNA cloning was present at a frequency of < 0.1, and was not studied further.

We quantified the relative frequency of each major allele in F2 hybrids using a frequency index equal to frequency in human-preferring F2s minus frequency in guinea-pig-preferring F2s divided by the sum of the frequencies in the two types of F2s.

Confirmation of *Or4* as a single copy gene. The *Ae. aegypti* reference genome contains another olfactory receptor, named *Or5*, that is 96% identical to the *Or4* gene across the coding sequence, leading to a predicted Or5 protein that is 97% identical to the *Or4* protein. Although we did not sample any sequences identical to *Or5* in our field-derived colonies, some of the alleles we sampled were more similar to *Or5* than to *Or4*, raising the possibility that they actually belong to this second hypothetical locus. To confirm that all sampled alleles segregate at a single locus, we reanalysed our PacBio genotyping data, focusing on the number of alleles carried by individual mosquitoes. As described above, we amplified a diagnostic segment of *Or4/Or5* from the genomic DNA of 270 females and sequenced ~150 pieces of DNA from each resulting amplicon. The reads for each mosquito were then assigned to alleles based on sequence. The vast majority of reads for each mosquito corresponded to a single allele (homozygotes) or were evenly split between two alleles (heterozygotes) (Extended Data Fig. 4). No mosquito had substantial numbers of reads assigned to more than two alleles. We conclude that *Or4* is a single locus

with multiple diverse alleles and suspect that the gene annotated as *Or5* represents a misassembly of *Or4* allelic sequences segregating within the genome reference strain. **Or4 transmembrane prediction and snake plots.** For Extended Data Fig. 3, we used TMHMM (v.2.0)⁶⁴ to predict the location of transmembrane domains in the A reference allele of *Or4*. The output was manually edited to remove a predicted transmembrane domain that occurred in an anomalous position relative to predictions for the olfactory co-receptor Orco and other ligand-selective ORs. Snake plots were generated using TOPO2 (ref. 65).

Single sensillum electrophysiology. We used single sensillum recording (SSR) to characterize the spontaneous activity and odour-evoked responses of 7 major *Or4* alleles heterologously expressed in *Drosophila* ab3A olfactory sensory neurons. Recordings were conducted as described previously⁴¹ using 5–11-day-old females and a modified method for immobilizing the third antennal segment. We glued a small wedge-shaped plastic Lego piece (rectangular footprint 15 mm × 10 mm with angled surface rising at 45° to 10 mm above base) to a glass microscope slide such that a coverslip affixed to its upper face with double stick tape would protrude upwards at a fixed angle of 45°. We then immobilised the fly in a pipette tip mounted on a ball of dental wax opposite the coverslip as described⁶⁶. After bringing the base of the fly's antennae into contact with the edge of the coverslip, we used two small drops of UV glue (KOA 300, Kexmert) to secure the rim of the cut pipette tip on either side of the protruding fly head to the edge of the coverslip and cured it for 10 s with a 405 nm violet laser pointer. We then used a glass micropipette to gently lower one antenna towards the angled coverslip until the arista contacted a small drop of UV glue placed directly below, and immediately cured the glue with the laser. These modifications prevented the fly preparation and antenna from drifting away from the coverslip over the course of a recording. Neither transgenic nor wild type ab3 sensilla responded to the odour of the glue.

We prepared serial dilutions of sulcatone (C.A.S. 409-02-9, Sigma-Aldrich) v/v in paraffin oil and loaded 30 µl aliquots into odour delivery pipettes on the day of recording as described⁶⁶. During recording, we applied a continuous stream of charcoal-filtered air to the insect antenna. For each stimulus, we first cleared the odour delivery pipette of accumulated volatiles by redirecting a fraction of the air stream through it and away from the preparation for a 1 s pulse. Fifteen seconds later, we then delivered the stimulus by redirecting air through the pipette for another 1 s pulse, but this time with the tip inserted back into the air stream flowing over the antenna. We used each pipette no more than twice.

We applied dilutions to a single ab3 sensillum per fly in the following order: solvent, 10^{-7} , 10^{-6} , 10^{-5} , 10^{-4} , 10^{-3} , 10^{-2} . To increase resolution for the most biologically relevant doses, we sometimes stimulated a second sensillum on the same fly with 10^{-5} and 10^{-4} dilutions only. Final sample sizes were $n = 13$ –17 sensilla per genotype for 10^{-5} and 10^{-4} and $n = 8$ –10 sensilla per genotype for solvent and all other dilutions. We identified ab3 sensilla by size, location on the antenna, lack of A cell response to the ab2A ligand ethyl acetate (10^{-2}) (C.A.S. 141-78-6, Sigma-Aldrich), and strong B cell response to its cognate ligand, 2-heptanone (10^{-4})⁶⁷ (C.A.S. 110-43-0, Sigma-Aldrich). We confirmed the *Dhalol/Dhalo* genotype of each fly by lack of A cell response to ethyl hexanoate (10^{-4}) (C.A.S. 123-66-0, Sigma-Aldrich).

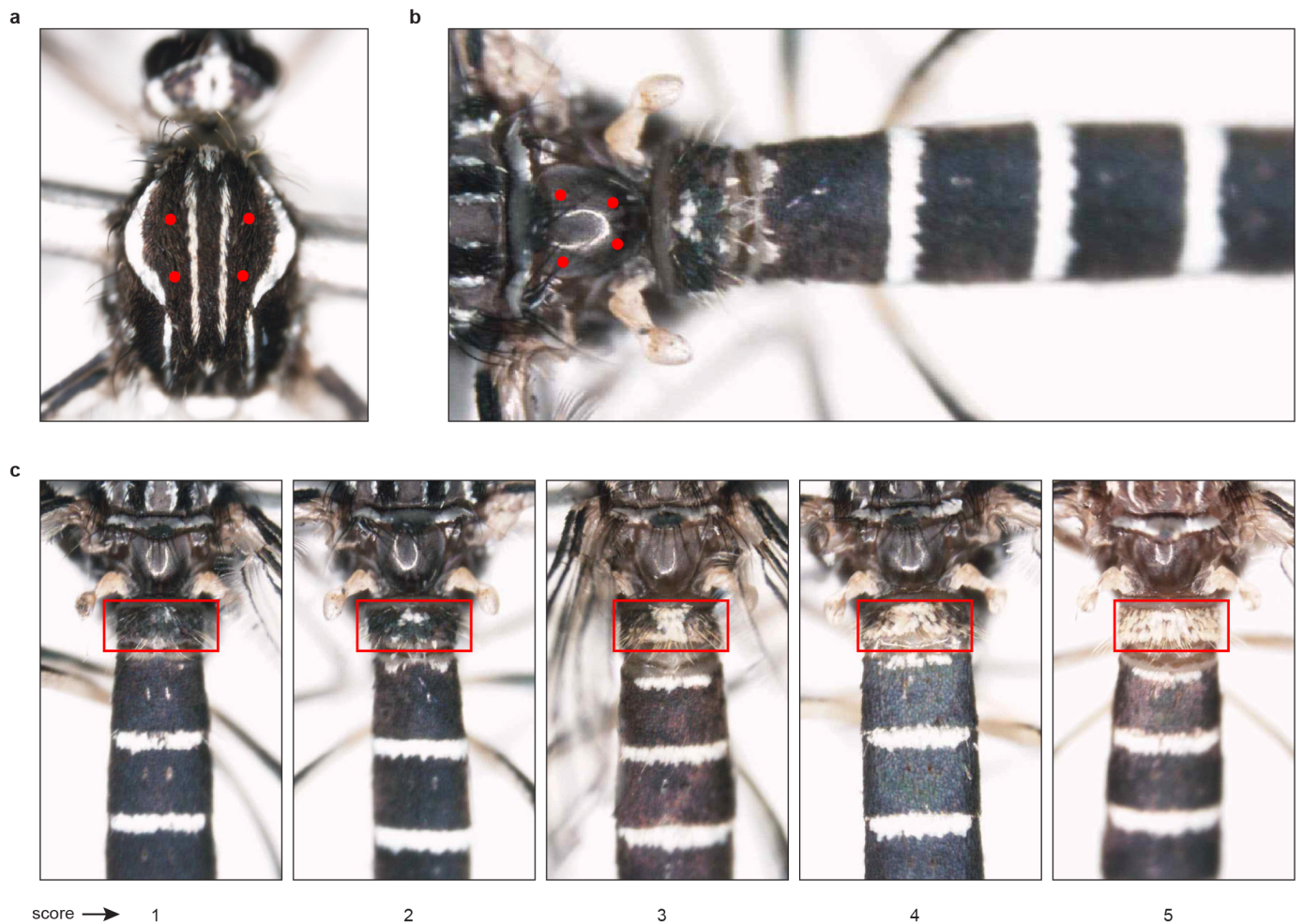
We recorded and processed neuronal activity using AutoSpike software (Syntech). Spontaneous activity was averaged over a 14 s period for each sensillum. Evoked response was calculated by subtracting average spontaneous activity in the 3 s before stimulus onset from average activity in the 1 s after stimulus onset. The distance air travels through the odour delivery system causes a delay between digitally recorded onset and the time the stimulus reaches the antenna. We therefore defined stimulus onset on the basis of the excitatory response elicited in the ab3B cell by 10^{-2} sulcatone; this invariably occurred at 200 ms after the computer-recorded onset. We compared spontaneous and evoked activity across fly genotypes using 1-way ANOVA followed by Tukey's test. We fit monotone cubic splines to evoked response data for each genotype (Fig. 5f) using the R function `smooth.monotone`.

Analysis of allele-specific expression. We estimated the allele-specific expression of the seven major *Or4* alleles in F2 hybrid RNA-seq libraries as follows. First,

we altered our reference transcriptome by removing reference transcripts for *Or4* and *Or5* and adding separate transcript sequences for each major *Or4* allele A–G. We then realigned unfiltered RNA-seq reads to this transcriptome using Bowtie2 v.2.1.0 (ref. 68) and estimated expression using eXpress v.1.5 (ref. 69) specifying *Or4* alleles as haplotypes of a single locus. We used the Bowtie2 alignment parameters recommended for use with eXpress, allowing an unlimited number of hits per read (`-a -rdg 6,5 -rfg 6,5 -score-min l,-0.6,-0.4`). eXpress then uses a probabilistic model to weight the hits from which it may be derived⁶⁹. The overall alignment rate was 47–51% for all libraries. For each F2 pool, we estimated expression for the two replicate libraries separately and then averaged them. The replicate estimates for a single pool were generally closer to each other than either was to estimates from the other pool. Expression values for each allele are partly a function of the frequency of that allele among the 110–125 individuals whose antennae were dissected for RNA extraction. We therefore normalized each allele's expression in a given F2 pool by its frequency among the individuals that made up that pool to generate final estimates of allele-specific expression. We used a two-way ANOVA to test for significant differences between the estimates derived from human-preferring *versus* guinea-pig-preferring pools and among the estimates for individual alleles.

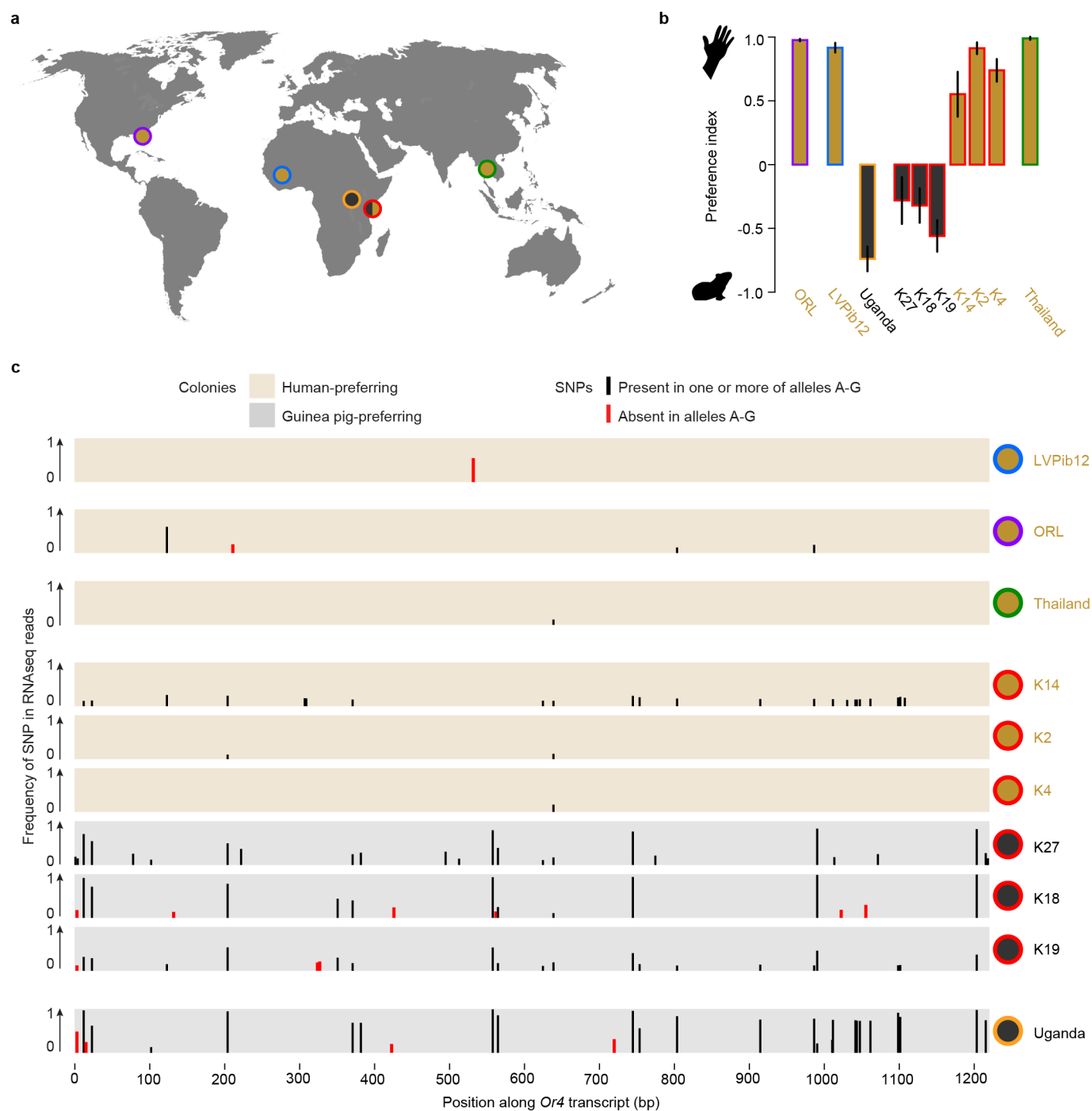
Statistical analysis. All statistical analyses were performed with JMP software v. 8 (SAS Institute, Inc.) or R software v.2.15.0 (<http://www.r-project.org/>). For all parametric tests, including *t*-tests and analyses of variance, data were tested and met the assumptions of normality and homogeneity of variance.

51. McClelland, G. A. H. A preliminary study of genetics and abdominal colour variations in *Aedes aegypti* (L.). *Ann. Trop. Med. Parasitol.* **54**, 305–320 (1960).
52. McClelland, G. A. H. A worldwide survey of variation in scale pattern of abdominal tergum of *Aedes aegypti* (L) (Diptera: Culicidae). *Trans. R. Entomol. Soc.* **126**, 239–259 (1974).
53. Mukwaya, L. Genetic control of feeding preferences in the mosquitoes *Aedes* (*Stegomyia*) *simpsoni* and *aegypti*. *Physiol. Entomol.* **2**, 133–145 (1977).
54. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
55. Kent, L. B., Walden, K. K. O. & Robertson, H. M. The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*. *Chem. Senses* **33**, 79–93 (2008).
56. Zhou, J.-J., He, X., Pickett, J. & Field, L. Identification of odorant-binding proteins of the yellow fever mosquito *Aedes aegypti*: genome annotation and comparative analyses. *Insect Mol. Biol.* **17**, 147–163 (2008).
57. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnol.* **31**, 46–53 (2013).
58. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
59. Hallem, E. A., Ho, M. & Carlson, J. The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117**, 965–979 (2004).
60. Harraca, V., Ryne, C., Birgersson, G. & Ignell, R. Smelling your way to food: can bed bugs use our odour? *J. Exp. Biol.* **215**, 623–629 (2012).
61. Stensmyr, M. C., Giordano, E., Balloi, A., Angioy, A. M. & Hansson, B. S. Novel natural ligands for *Drosophila* olfactory receptor neurones. *J. Exp. Biol.* **206**, 715–724 (2003).
62. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
63. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
64. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
65. Johns, S. J. TOPO2, transmembrane protein display software. <http://www.sacs.ucsf.edu/TOPO2/>.
66. Pellegrino, M., Nakagawa, T. & Vosshall, L. B. Single sensillum recordings in the insects *Drosophila melanogaster* and *Anopheles gambiae*. *J. Vis. Exp.* **36**, 1–5 (2010).
67. Hallem, E. A. & Carlson, J. R. Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).
68. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
69. Roberts, A. & Pachter, L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods* **10**, 71–73 (2013).



Extended Data Figure 1 | Measuring colour and scaling of adult female *Ae. aegypti* mosquitoes. **a**, Representative photograph used to measure scale colour (Fig. 1e, g). Red dots mark the approximate position of 4 points where the colour of dark scales on the scutum was assessed. **b**, Representative photograph used to measure cuticle colour (Fig. 1f, h). Red dots mark the

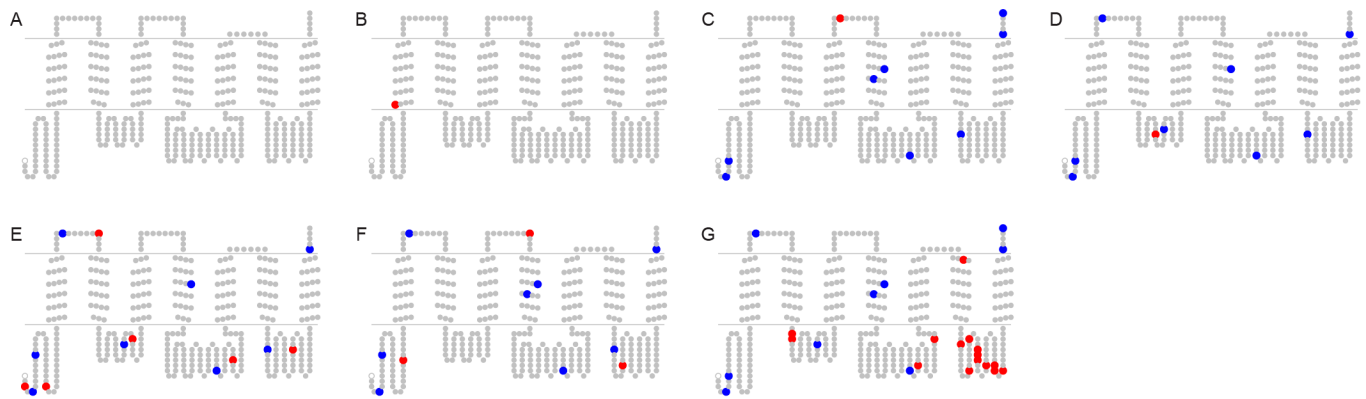
approximate position of 4 points where the colour of bare cuticle on the circular postnotum was assessed. **c**, Representative photographs used to assess the extent of white scaling on the first abdominal tergite (Fig. 1i), outlined with the red rectangle. Each individual is representative of the scaling score shown at the bottom.



Extended Data Figure 2 | Or4 coding sequence variation in human-preferring and guinea-pig-preferring colonies from around the world.

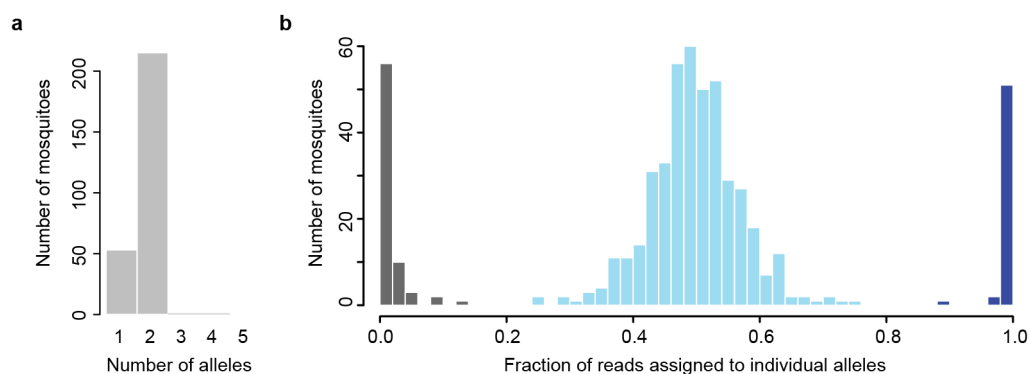
a, Geographical origin of colonies characterized in **b** and **c**. Circle fill colour indicates preference of strains. Circle outline colour indicates origin: Purple, laboratory strain derived from USA; blue, reference genome strain derived from West Africa; orange, Uganda; red, Kenya, green, Thailand. **b**, Host preference assayed in the live host olfactometer. Data for Thailand, K14, K2, K4,

K27, K18, K19, and Uganda are reprinted from Fig. 2g. **c**, Frequency of non-synonymous single nucleotide polymorphisms (SNPs) in female antennal RNA-seq reads. SNPs are defined as differences from the A reference allele. SNPs with frequency ≤ 0.1 are not shown. Vertical black and red lines indicate SNPs that were present and absent, respectively, in the major alleles subject to functional analysis.



Extended Data Figure 3 | Amino acid differences of major Or4 protein alleles. Dots represent amino acid differences with respect to the genome reference, not an inferred ancestor. Red dots indicate differences that are unique to the given allele. Blue dots indicate differences that are shared among

multiple alleles. Snake plots are based on the predicted orientation and location of transmembrane domains. Extracellular loops are oriented up and cytoplasmic loops are oriented down. Allele names are indicated to the left of each snake plot.

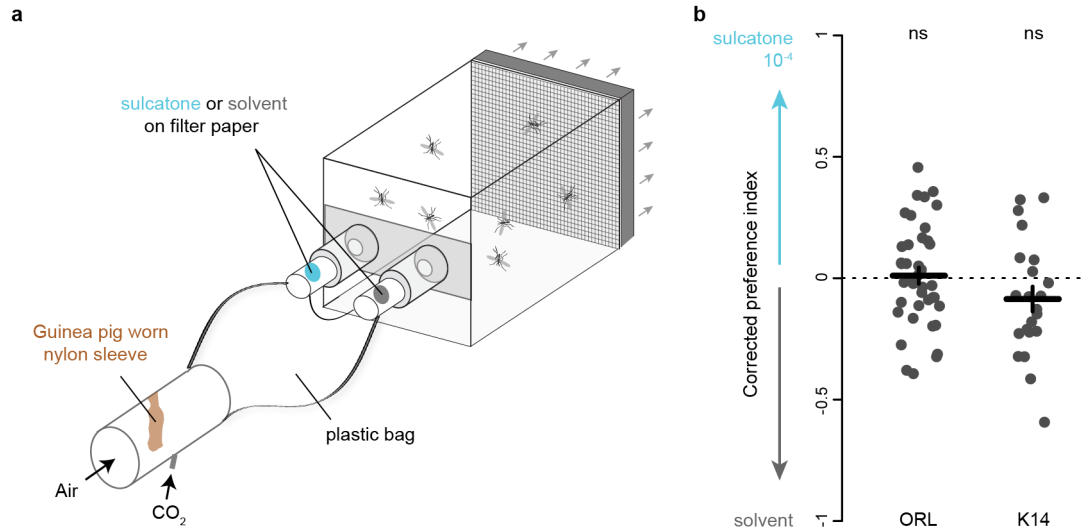


Extended Data Figure 4 | Evidence that *Or4* is a single copy gene.

a, Histogram showing the number of alleles represented in the *Or4*-derived PacBio reads obtained for each of 270 parent and F2 hybrid mosquitoes. Alleles were only considered if they received at least 5% of an individual's reads.

b, Histogram showing the fraction of reads from individual mosquitoes assigned to individual alleles. For all 270 mosquitoes, individual alleles were

represented by either very few reads (grey bars, inferred to result from allele or barcode assignment errors or polymerase chain reaction contaminants), approximately half the reads (light blue bars, inferred to represent the two alleles in heterozygotes), or over 98% of all reads (dark blue bars, inferred to represent the single allele carried by homozygotes).



Extended Data Figure 5 | Response of human-preferring mosquitoes to sulcatone-scented guinea-pig odour. **a**, Olfactometer apparatus in which 50 mosquitoes per trial were given a choice between guinea-pig odour/CO₂ mix supplemented with solvent on one side and sulcatone 10⁻⁴ on the other side. **b**, Corrected preference for sulcatone vs solvent ports is indicated. Data were corrected for the daily average left–right side bias observed across

2–3 solvent vs solvent tests conducted on each day of testing. An index value of 1 indicates strong preference for the sulcatone side, whereas –1 indicates strong preference for the solvent side. Neither mosquito colony showed a preference significantly different from zero (one-sample *t*-test $P = 0.76$ for ORL, $P = 0.11$ for K14). The trials for each colony were performed across 4–8 days ($n = 40$ for ORL and $n = 22$ for K14).

Clathrin regenerates synaptic vesicles from endosomes

Shigeki Watanabe¹, Thorsten Trimbuch², Marcial Camacho-Pérez², Benjamin R. Rost^{2,3}, Bettina Brokowski², Berit Söhl-Kielczynski², Annegret Felies², M. Wayne Davis¹, Christian Rosenmund² & Erik M. Jorgensen¹

Ultrafast endocytosis can retrieve a single, large endocytic vesicle as fast as 50–100 ms after synaptic vesicle fusion. However, the fate of the large endocytic vesicles is not known. Here we demonstrate that these vesicles transition to a synaptic endosome about one second after stimulation. The endosome is resolved into coated vesicles after 3 s, which in turn become small-diameter synaptic vesicles 5–6 s after stimulation. We disrupted clathrin function using RNA interference (RNAi) and found that clathrin is not required for ultrafast endocytosis but is required to generate synaptic vesicles from the endosome. Ultrafast endocytosis fails when actin polymerization is disrupted, or when neurons are stimulated at room temperature instead of physiological temperature. In the absence of ultrafast endocytosis, synaptic vesicles are retrieved directly from the plasma membrane by clathrin-mediated endocytosis. These results may explain discrepancies among published experiments concerning the role of clathrin in synaptic vesicle endocytosis.

Clathrin is thought to act at the plasma membrane of synapses to retrieve synaptic vesicle membrane and proteins. There is an extensive literature supporting this conclusion. Classic ultrastructural studies of frog neuromuscular junctions revealed the presence of clathrin coats on the plasma membrane after stimulation, suggesting that synaptic vesicles are reconstituted at the surface¹. Biochemical purification of clathrin-coated vesicles from rat brain demonstrated that synaptic vesicle proteins copurify with clathrin and AP2 (ref. 2), and AP2 and clathrin are sufficient to generate vesicles from purified brain lipids^{3,4}. Transmembrane vesicle proteins, such as synaptotagmin^{5–7} and synaptobrevin^{8,9}, interact with adaptor proteins at the plasma membrane and these interactions are required for regenerating functional synaptic vesicles. These studies lend strong support to the idea that clathrin acts at the plasma membrane to regenerate functional synaptic vesicles. However, recent morphological studies^{10,11} suggest that endocytosis may be much faster than previously described for clathrin-mediated endocytosis.

'Flash-and-freeze' fixation combines optogenetic stimulation with high-pressure freezing to capture events at synapses milliseconds after stimulation. Our ultrastructural studies demonstrated that endocytic pits, which lacked stereotypical clathrin coats, appeared ~50 ms after stimulation at both *Caenorhabditis elegans*¹¹ and mouse synapses¹⁰. Ultrafast endocytosis is compensatory, that is, membrane retrieval is triggered by the membrane added by fusion¹⁰. Under our stimulation conditions this appears to be due to multiquantal release¹⁰, as has been previously observed at cultured hippocampal synapses¹². The large endocytic vesicles retrieved by ultrafast endocytosis are too big to be functional synaptic vesicles—they are typically equivalent to the surface area of 4 synaptic vesicles—and the fate of these vesicles was unknown.

Here we examined events that occur up to 20 s after stimulation. We performed 'flash-and-freeze' experiments on mouse hippocampal neurons and analysed the synaptic ultrastructure blind from ~200 synaptic profiles per time point. A total of 10,514 synaptic profiles and 361 tomograms were analysed. We applied a single stimulus in the presence of 4 mM Ca^{2+} or 10 stimuli (20 Hz) in the presence of 2 mM Ca^{2+} . At physiological temperatures (34–37 °C), clathrin is not required for ultrafast endocytosis but is required to bud synaptic vesicles from synaptic

endosomes (Extended Data Fig. 1). However, at room temperature (~22 °C), clathrin functions at the plasma membrane during endocytosis.

Synaptic vesicles reform from endosomes

Clathrin polyhedral coats on endocytic membranes are distinctive in electron micrographs (Fig. 1 and Extended Data Fig. 2). In our preparations, these coats are observed on pits on the cell body (Extended Data Fig. 2d) but rarely observed on the plasma membrane at synapses. To test the requirement of clathrin after stimulation, a single stimulus was applied to hippocampal cells expressing a variant of channelrhodopsin (ChetaTC)¹³. The experiments were performed in the presence of 4 mM Ca^{2+} in the external solution at 34 °C (Fig. 1) and 37 °C (Extended Data Fig. 2). After a single stimulus, endocytic pits that lack distinctive clathrin coats formed 50–100 ms after fusion (Fig. 1a, left panel; Extended Data Fig. 2a, left panel)¹⁰. These invaginations resolved into large vesicles about 80 nm in diameter at the lateral edges of the active zone (Fig. 1a and Extended Data Fig. 2a). The number of large endocytic vesicles adjacent to the plasma membrane peaked at 100 ms (Fig. 1d; Extended Data Fig. 2e). Thereafter, endosome-like structures began accumulating in the bouton and peaked at 1 s (Fig. 1b, d and Extended Data Fig. 2b, e). These organelles are larger in diameter (116.4 ± 2.5 nm, 8 synaptic vesicle equivalents) than the endocytic vesicles at the periphery (80.6 ± 0.7 nm; 4 synaptic vesicle equivalents; Extended Data Fig. 3a, see Methods), suggesting that the large endocytic vesicles fuse to form a synaptic endosome¹⁴. Tomographic reconstructions of these endosomes demonstrated that they are not connected to the plasma membrane (Extended Data Fig. 2c and Supplementary Video 1). Clathrin-like coats were visible on some of these endosomes (Fig. 1b, c and Extended Data Fig. 2b–d), and budded endosomes peaked 3 s after stimulation (Fig. 1c, d and Extended Data Fig. 2b, c, e). The decline in endosomes was accompanied by an accumulation of clathrin-coated vesicles inside the bouton at 3 s (Fig. 1d and Extended Data Fig. 2e, f). By contrast, clathrin-coated pits on the plasma membrane were only observed in 0.4% of synaptic profiles (4/907) between 1 to 10 s after stimulation (1 s (2/332); 3 s (2/345); 10 s (0/330)) (Fig. 1d and Extended Data Fig. 2e). These results suggest that clathrin does not

¹Department of Biology and Howard Hughes Medical Institute, University of Utah, Salt Lake City, Utah 84112-0840, USA. ²Neuroscience Research Center Charité Universitätsmedizin Berlin, Berlin 10117, Germany. ³German Center for Neurodegenerative Diseases (DZNE), 10117 Berlin, Germany.

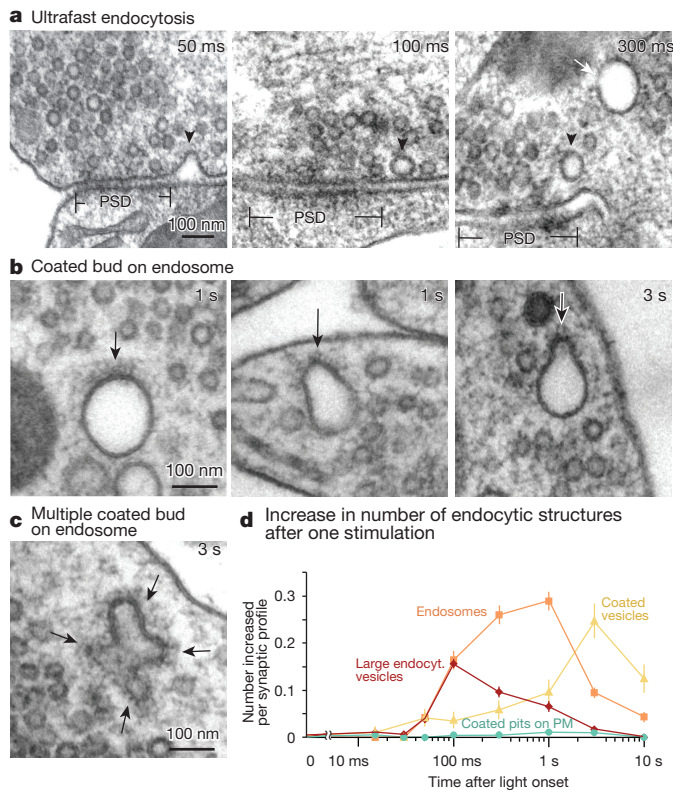


Figure 1 | Synaptic vesicles are regenerated from endosomes at 34 °C. Hippocampal synapses were stimulated once and frozen at the indicated times. **a**, Electron micrographs showing invaginations and large endocytic vesicles (black arrowheads) recovered via ultrafast endocytosis. White arrow, an endosome in the centre of the bouton. PSD, postsynaptic density. **b**, **c**, Micrographs showing single coated buds (**b**) and multiple coated buds (**c**) forming on an endosome. Black arrow, coated bud on an endosome. **d**, Increase in the number of each endocytic structure per synaptic profile after a single stimulus. Approximately 60% of synapses had endosomes in the unstimulated control; this baseline value was subtracted from endosome numbers. After stimulation, two endosomes were observed in 30% of the synapses. The prevalence of large endocytic vesicles and endosomes is followed by an increase in the number of clathrin-coated vesicles. Coated pits were rarely observed on the plasma membrane (PM). The standard error of the mean is shown in each graph. For *n* values, detailed numbers and statistical analysis, see Supplementary Table 1. Similar results were obtained from two independent cultures, and at least 178 synaptic profiles were analysed per time point (Supplementary Table 1). Endosomes are defined as vesicles larger than ~100 nm and greater than 50 nm from the active zone (see Methods and Extended Data Fig. 3).

regenerate synaptic vesicles via endocytosis at the plasma membrane but rather by budding vesicles from an endosome.

Clathrin not required for endocytosis

To test the role of clathrin in ultrafast endocytosis, we reduced expression of clathrin using RNAi¹⁵. Cultures were infected with lentivirus expressing a short hairpin RNA (shRNA) targeted against clathrin heavy chain mRNA, or a scrambled shRNA control. Clathrin was significantly reduced seven days after infection as determined by western blot (~80% reduction, *n* = 3; Extended Data Fig. 4a) and immunostaining (64% reduction; Extended Data Fig. 4b, d, e). Similarly, transferrin uptake was reduced by 66% (Extended Data Fig. 4c, f). Clathrin knockdown reduced release-ready vesicles but did not affect the exocytic machinery in electrophysiological recordings (Fig. 2a, Extended Data Fig. 5a–f, see Supplementary Information). Likewise, a smaller docked pool of vesicles was observed by electron microscopy, but the overall morphology of synapses was normal (Extended Data Fig. 5g, h, see Supplementary Information).

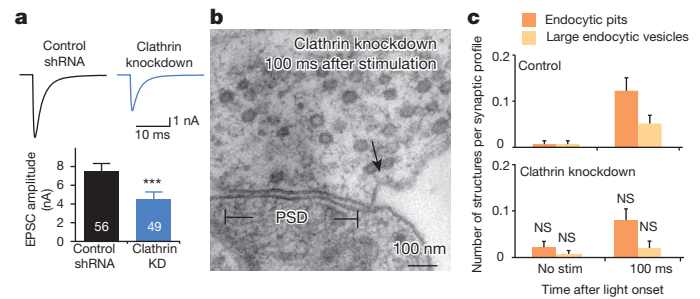


Figure 2 | Ultrafast endocytosis is clathrin-independent. **a**, Top, average traces for excitatory post-synaptic currents (EPSCs) in a control and a clathrin knockdown neuron from autaptic cultures. Bottom, mean amplitude of EPSCs. The amplitude is reduced by 41% in the knockdown cells. **b**, An electron micrograph of a synapse frozen 100 ms after stimulation. Ultrafast endocytic pit (arrow) is observed flanking the active zone. **c**, Average number of endocytic pits and large endocytic vesicles in control and clathrin knockdown neurons with or without stimulation (control 100 ms: 0.17 ± 0.03 endocytic structures per profile; knockdown 100 ms: 0.10 ± 0.03 endocytic structures per profile, $P = 0.06$). The *P* values are calculated against the matched time points in the control shRNA treated cells. *** $P < 0.001$. NS, not significant. The standard error of the mean is shown in each graph. For *n* values, detailed numbers and statistical analysis, see Supplementary Table 1.

Clathrin knockdown did not disrupt ultrafast endocytosis, which remained proportional to vesicle fusion. Following a single stimulus (34 °C, 4 mM Ca^{2+}), endocytic pits and large endocytic vesicles appeared at the edges of the active zone both in the control and clathrin knockdown cells (Fig. 2b, c). The number of endocytic pits and vesicles was decreased by 41% (Fig. 2c), closely matching the reduction in exocytosis of synaptic vesicles (Extended Data Fig. 5h). These results suggest that membrane retrieval via ultrafast endocytosis does not depend on clathrin, which is consistent with the lack of stereotypic clathrin coats on pits during ultrafast endocytosis (Fig. 1a and Extended Data Fig. 2a).

Clathrin regenerates vesicles at endosomes

To determine whether clathrin is required to resolve endosomes into synaptic vesicles, control and clathrin knockdown cells were stimulated in the presence of cationized ferritin (4 mM Ca^{2+} , 34 °C). Cationized ferritin adheres to membranes and serves as a fluid phase marker for endocytic trafficking. In control cultures 1 s after a single stimulus (Fig. 3a, b), ferritin molecules were found in endosomes but not in clathrin-coated vesicles. At 3 s, endosomes were budding and ferritin began to appear in clathrin-coated vesicles. The size of endosomes was larger than endocytic vesicles, again suggesting that endocytic vesicles carrying ferritin fuse to form synaptic endosomes (Extended Data Fig. 3b, c). Sixteen ferritin-positive endosomes were reconstructed from tomograms, none of them extended a tubule to the plasma membrane (Extended Data Fig. 6a). However, half of those endosomes were not fully contained within the 200 nm tomogram. Therefore we reconstructed 11 complete synapses by assembling serial tomograms of synapses 3 s after stimulation. None of the 17 complete end-to-end endosomes were connected to the plasma membrane (6 of which contained ferritin), suggesting that they are true intracellular organelles, not extensions of the plasma membrane (Supplementary Videos 1–5). The decline in the number of endosomes is followed by an accumulation of ferritin-positive coated vesicles and synaptic vesicles (Fig. 3b). Some clathrin-coated vesicles were observed before 3 s after stimulation (1 s, 4/149 synaptic profiles and Fig. 1d), but they did not contain ferritin molecules, suggesting that they were derived from pre-existing endosomes. The total number of ferritin granules per synaptic profile did not increase, suggesting that an additional wave of endocytic events did not occur during those 20 s (Extended Data Fig. 6c). Some of the ferritin-filled small vesicles docked to the active zone (Fig. 3a, right panel; Extended Data Fig. 6b, d), suggesting that these vesicles are synaptic vesicles. These results indicate that recently endocytosed membrane passes through endosomes to regenerate synaptic vesicles within about 5–6 s after fusion.

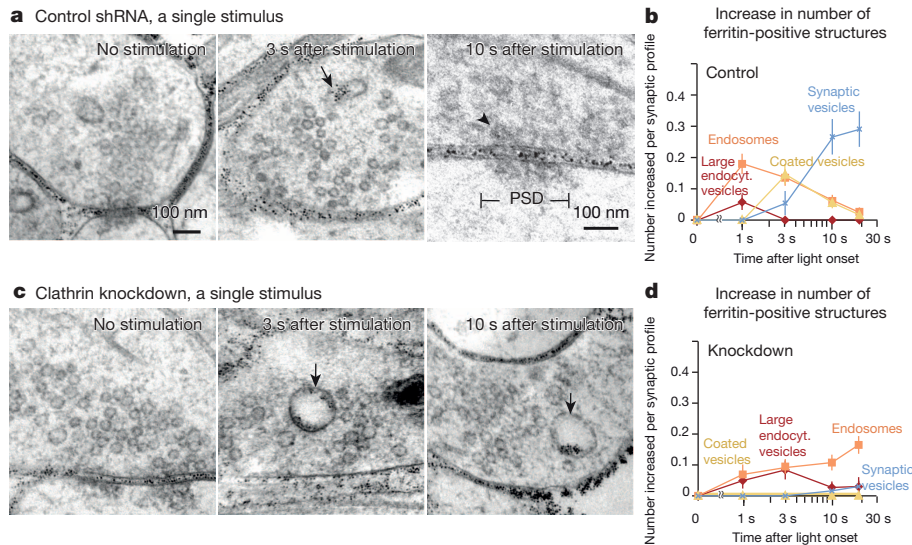


Figure 3 | Following a single stimulus, clathrin is required at endosomes to regenerate synaptic vesicles. **a, c**, Electron micrographs showing ferritin uptake in control (**a**) and clathrin knockdown (**c**) synapses at different time points after stimulation. In control neurons, ferritin is observed in large endocytic vesicles after stimulation (middle) and in synaptic vesicles (right), but it is trapped in endosomes in the clathrin knockdown cells. Black arrows indicate ferritin-positive structures. The black arrowhead in **a** represents ferritin-positive synaptic vesicles docked in the active zone. **b, d**, Average increase in ferritin-positive endocytic structures in all profiles. **b**, In control cells, the total number of synapses containing ferritin remained at 26% at the 3, 10, and 20 s time points; suggesting that 74% of synapses were silent.

In clathrin knockdown cultures, ferritin-filled endosomes increased after 1 s (Fig. 3d). However, these endosomes did not decline across the 3 s, 10 s or 20 s (Fig. 3c, d). These endosomes formed almost no coated buds and remained spherical (Fig. 3c), suggesting that membrane curvature of the buds requires clathrin. Together, these results suggest that clathrin is required at endosomes to regenerate synaptic vesicles following a single stimulus.

Clathrin at endosomes after 10 stimuli

At mammalian central synapses, action potentials often fire in bursts^{16,17} and could exhaust ultrafast endocytosis. To assay endocytosis after high-frequency stimulation, we delivered 10 stimuli (20 Hz for 500 ms in 2 mM Ca^{2+} , 34 °C) and froze cultures 1 s, 3 s, 10 s, and 20 s after the end of the stimulus. Under these conditions, 90% of the cells fired action potentials for all 10 light pulses (Extended Data Fig. 7a). Using cationized ferritin, we followed the fate of the newly endocytosed vesicles (Fig. 4 and Extended Data Fig. 3d, e). The results were similar to those after a single stimulus. Coated pits were not observed on the plasma membrane. Ferritin accumulation in endosomes peaked at 1 s; endosomes exhibited coated buds at 3 s, accompanied by the appearance of clathrin-coated vesicles and some uncoated synaptic vesicles. All vesicles were uncoated by 10 s (Fig. 4a, b and Extended Data Fig. 7b). Ferritin-filled vesicles were eventually recruited to the active zone, suggesting that these vesicles are probably synaptic vesicles (Extended Data Fig. 7c).

We then applied an identical 20 Hz stimulation to cultures treated with clathrin shRNA. In these cultures ferritin was taken up into large endosomes, again suggesting that membrane internalization after a high frequency burst does not require clathrin (Fig. 4c). However, ferritin-containing endosomes were abundant and were not resolved into vesicles even after 20 s (Fig. 4c, d), indicating that clathrin is essential for regenerating vesicles from endosomes after a burst of action potentials. The acute inhibition of clathrin function by Pitstop 2, a potent inhibitor of clathrin-mediated endocytosis, (30 μM , 2 min) also blocked regeneration of synaptic vesicles from endosomes (Extended Data Fig. 8; see

The number of synaptic vesicles is higher because after an endosome is resolved it leads to ~2 vesicles that contain a clump of ferritin. The number of clathrin-coated vesicles is less than expected given that endosomes contain 5–8 synaptic vesicles worth of membrane. This discrepancy is probably because synaptic vesicles will be distributed among many thin sections. **d**, In the clathrin knockdown, ferritin stalled in endosomes and did not progress into synaptic vesicles. Clathrin-coated pits were not present on the plasma membrane at any time point and are therefore not plotted. The standard error of the mean is shown in each graph. Experiments are from two independent cultures and at least 149 synaptic profiles were analysed per time point (Supplementary Table 1).

Supplementary Information). These results suggest that following a short burst of action potentials, clathrin acts on endosomes to regenerate synaptic vesicles.

To determine whether intense stimulation triggers clathrin-mediated endocytosis, we applied trains of 3, 10, 30, and 100 light pulses at 20 Hz to hippocampal neurons and froze them 10 s after stimulation (2 mM Ca^{2+} , 34 °C). To arrest all endocytic intermediates at the plasma membrane, we applied 80 μM dynasore for 30 s before the stimulation. Dynasore is a drug that binds dynamin¹⁸, although it also has off-target effects¹⁹, it nevertheless effectively locks both clathrin-mediated endocytosis and ultrafast endocytosis at a late stage of vesicle scission^{10,18}. After dynasore treatment, no clathrin-coated pits were trapped on the plasma membrane under any condition. Instead, strings of uncoated vesicles attached to the plasma membrane accumulated at the edges of active zones (Fig. 4e, f). The diameter of vesicles in strings (53.3 ± 3.7 nm) was larger than synaptic vesicles (~40 nm) but smaller than large endocytic vesicles formed after a single stimulus (~80 nm), reflecting reduced fusion during repetitive stimulation. These results suggest ultrafast endocytosis is responsible for removing membrane from the surface even during high frequency bursts of action potentials and that clathrin functions on endosomes rather than at the plasma membrane to regenerate synaptic vesicles.

Ultrafast pathway is temperature sensitive

Actin polymerization is required for ultrafast endocytosis¹⁰. To study synaptic vesicle recycling in the absence of ultrafast endocytosis, we blocked actin polymerization. We treated cultured neurons with 10 μM latrunculin-A for 30 s, stimulated once by a 10 ms light flash, and frozen 3 or 10 s later (4 mM Ca^{2+} , 34 °C). Large endocytic vesicles adjacent to the active zone were absent following the latrunculin-A application, indicating that ultrafast endocytosis was blocked (Fig. 5a–d). Under these conditions, clathrin-coated pits were observed on the plasma membrane 3 s after stimulation, followed by an accumulation of coated vesicles between 3 and 10 s, and the appearance of large vesicles at 10 s (Fig. 5c, d). The size of coated pits, coated vesicles and synaptic vesicles were similar (Extended Data Fig. 9a).

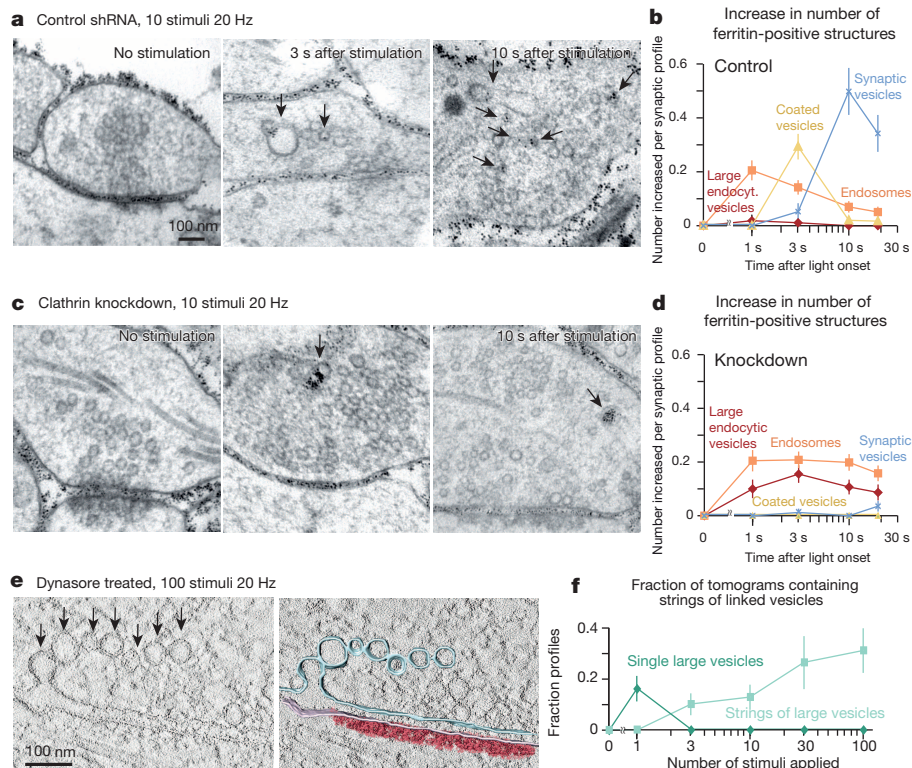


Figure 4 | Following high-frequency stimulation, clathrin is required at endosomes to regenerate synaptic vesicles. **a, c, e**, Electron micrographs showing ferritin uptake in control (**a**) and clathrin knockdown neurons (**c**) at different time points after stimulation. Black arrows indicate ferritin-positive structures. **b, d**, Average number of ferritin-positive endocytic structures increased compared to unstimulated cells (0 ms time point) infected with control shRNA (**b**) or clathrin knockdown shRNA (**d**). Clathrin-coated pits were not present on the plasma membrane at any time points and thus not plotted. In the controls, synapses which exhibited ferritin uptake remained at 32%, but the number of ferritin structures per profile increased because endosomes were typically resolved into ~2 ferritin-positive vesicles. Experiments are from two independent cultures and at least 123 synaptic profiles were analyzed per time point (Supplementary Table 1). **e**, Virtual section from an electron tomogram (left) and reconstruction (right) showing a string of large endocytic vesicles trapped on the membrane by dynasore

treatment following 100 stimuli at 20 Hz. Multiple large endocytic vesicles (black arrows) are attached to one another and remain connected to the plasma membrane. **f**, Fraction of tomograms that contain vesicle strings attached to the plasma membrane following 1, 3, 10, 30, and 100 stimuli (20 Hz). Tomograms from 100-nm thick sections were reconstructed for each time point; only 1 vesicle string appeared in a given terminal. The number of tomograms with a vesicle string reached 31% following 100 stimuli, suggesting that ~70% of synapses are silent in these preparations. The number of vesicles on a string increases with repetitive stimulation: 1 stimulus (no vesicle string); 3 stimuli (2 vesicles per string); 10 stimuli (2.3 vesicles per string); 30 stimuli (5.6 vesicles per string); and 100 stimuli (4.8 vesicles per string). However, these are probably underestimates because strings that extend into neighbouring sections are not captured in these tomograms. The standard error of the mean is shown in each graph. For *n* values, detailed numbers and statistical analysis, see Supplementary Table 1.

Actin polymerization is known to be less efficient at room temperature in mammalian cells^{20,21}. To test whether ultrafast endocytosis fails at low temperature, we incubated cells at room temperature (22 °C) for 5 min before flash-and-freeze (Fig. 5e–h). In control cells maintained

at 34 °C, coated pits were not observed after stimulation (Fig. 5e–h, unstimulated, 0/251 synaptic profiles; 3 s, 0/248 synaptic profiles). However, at 22 °C, clathrin-coated pits appeared on the plasma membrane adjacent to the active zone 3 s after a single stimulus (Fig. 5g, h).

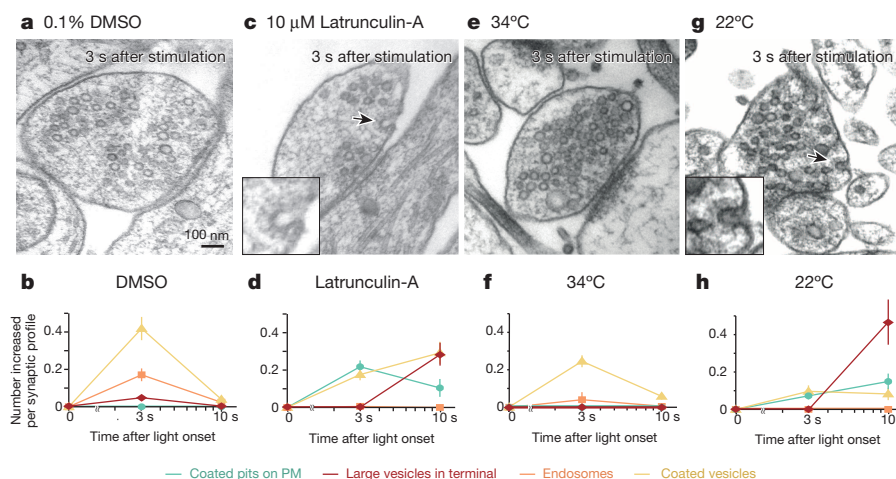


Figure 5 | Clathrin regenerates synaptic vesicles from plasma membrane in the absence of ultrafast endocytosis. **a, c, e, g**, Electron micrographs showing a synaptic terminal 3 s after a single stimulus from cells incubated with 0.1% dimethylsulphoxide (DMSO) (**a**), with 10 μM latrunculin-A (**c**), at 34 °C (**e**), and at 22 °C (**g**). Black arrows indicate clathrin-coated pits. **b, d, f, h**, Average number of endocytic structures per profile in cells incubated with 0.1% DMSO (**b**), 10 μM latrunculin-A (**d**), at 34 °C (**f**) and at 22 °C (**h**). The total number of structures per profile is plotted. Large vesicles accumulated in the centre of the bouton, probably reflecting the fusion of small endocytic vesicles to endosomes. The standard error of the mean is shown in each graph. Experiments are from two independent cultures and at least 83 synaptic profiles were analysed per time point (Supplementary Table 1).

Like in the latrunculin-A treated cells, large vesicles, possibly endosomes, accumulated in the centre of the bouton 10 s after stimulation^{22,23}, as observed previously in hippocampal cultures stimulated at room temperature^{22,23}. The diameter of coated pits and synaptic vesicles were similar (Extended Data Fig. 9b). These results suggest that failure of ultrafast endocytosis leads to clathrin-mediated endocytosis on the plasma membrane.

Temperature

In classic electron microscopy studies of stimulated synapses, clathrin-mediated endocytosis occurs about 20 s after stimulation^{24,25}. By contrast, clathrin-independent ultrafast endocytosis removes excess membrane 30–100 ms after stimulation^{10,11}, and as demonstrated here, synaptic vesicles are regenerated via an endosome 5–6 s after stimulation (Extended Data Fig. 1). These contradictory results can be reconciled simply by considering temperature.

Ultrafast endocytosis is only observed at physiological temperatures. When neurons are cooled to 22 °C, ultrafast endocytosis fails, and the slower clathrin-mediated process takes over. Actin polymerization is greatly reduced in cultured cells at non-physiological temperature^{26,27}. Because ultrafast endocytosis requires actin polymerization, it is probable that vesicle components accumulate on the plasma membrane at room temperature, and the clathrin machinery is recruited to the plasma membrane instead of the synaptic endosome. The shift to clathrin-mediated endocytosis may contribute to rapid synaptic depression that is observed at room temperature^{28,29}, perhaps because excess membrane in the active zone could block exocytosis.

Previous ultrastructural data have suggested the presence of a fast form of endocytosis at physiological temperature. When frog pectoral muscles were stimulated at 10 °C in the original Heuser and Reese experiments, clathrin-coated pits accumulated on the plasma membrane; whereas when stimulated at room temperature, two forms of endocytosis were observed: a fast form 1 s after stimulation, and a slow form 20 s after stimulation²⁴. At snake neuromuscular junctions, clathrin-coated pits accumulated when stimulated at 7 °C³⁰, whereas at room temperature, membrane was retrieved in 1–2 s³¹. In rat hippocampal neurons, endosomes (> 70 nm) were observed after stimulation at physiological temperature but not at room temperature³². Capacitance studies also indicate fast and slow mechanisms. In the calyx of Held, fast endocytosis was observed at physiological temperature (35–37 °C), but was abolished at room temperature³³. Fast endocytosis was also observed in fish retinal bipolar cells, membrane was recovered ~1 s after a single stimulation^{34,35}.

Ultrastructure and capacitance analyses only measure membrane endocytosis; protein endocytosis can be measured by pHluorin assays. At room temperature, a slow form of endocytosis is observed in hippocampal neurons with a surface dwell time of 15 s³⁶. When pHluorin experiments are conducted at physiological temperatures, the time constant for endocytosis is significantly faster: 14 s at 24 °C and 10 s at 30 °C, and as fast as 6 s at 37 °C^{37,38}. At first glance, a 6 s recovery for vesicle proteins agrees well with a 6 s recovery of morphologically defined synaptic vesicles. However, such a comparison is deceptive, as it is unlikely that proteins in a fusing vesicle could diffuse to the lateral edge of the active zone and into an endocytic vesicle within 50 ms. Nevertheless, temperature seems to be responsible for differences observed among ultrastructural studies.

Clathrin at endosomes

Endosomes are sorting compartments that receive cargo from transport vesicles generated by endocytosis from the plasma membrane. Heuser and Reese proposed in their original model that clathrin-mediated endocytosis generated a transport vesicle that fused to an endosome ('cisterna'), and synaptic vesicles were then regenerated from these endosomes¹. However, subsequent experiments demonstrated that these cisternae were likely to arise via bulk endocytosis due to non-physiological stimulation²⁵, and the model was amended to propose that synaptic vesicles

were generated directly from the plasma membrane via clathrin-mediated endocytosis, whereas bulk endocytosis removed excess membrane added during intense stimulation. Thus, the cisternae do not represent true endosomes as they do not arise via transport vesicles but rather from a simple pinching off from the plasma membrane. Moreover, it was observed that internal clathrin-coated buds were often, or even always, attached to the plasma membrane via long tubules^{1,4,39}, which led to the idea that the formation of clathrin-coated pits is limited to the plasma membrane. The different lipid compositions of endosomes and the plasma membrane support this notion. Endosomes are enriched for phosphatidylinositol 3-phosphate (PI(3)P), whereas the plasma membrane is enriched in phosphatidylinositol 4,5-bisphosphate (PI(4,5)P₂) and therefore can recruit clathrin-AP2 (ref. 40).

By contrast, our data suggest that clathrin regenerates synaptic vesicles from a synaptic endosome rather than from the surface. Tomographic reconstructions of endosomes failed to identify connecting tubules to the plasma membrane. Moreover, these compartments appear to be true endosomes, in that they are generated by transport vesicles. Because endosomes are twice as large as endocytic vesicles, the synaptic endosome probably forms by fusion of endocytic vesicles. The synaptic endosome is resolved into synaptic vesicles by clathrin-mediated budding, and for this reason it is unlikely that synaptic endosomes are typical endosomes. The presence of clathrin on these endosomes suggests they are enriched in PI(4,5)P₂. In this case, the membrane composition of 'synaptic endosomes' is more like the plasma membrane than that of classical endosomes.

A growing body of molecular studies support the conclusion that synaptic vesicles may be regenerated from endosomes. Live imaging in hippocampal neurons indicates that recently endocytosed vesicle proteins are sorted in an endosome after stimulation²². Endosomal Rab GTPases, such as Rab5 and Rab35, are abundant at the synapse and can be co-purified with synaptic vesicles⁴¹, and inhibition of Rab5 (ref. 42) and skywalker⁴³ (required for Rab35 function) results in an accumulation of endosomes in synaptic boutons. Finally, accumulation of endosomes is also observed in the absence of AP2, clathrin, clathrin accessory proteins or dynamin^{44–48}.

In summary, our results resolve an important contradiction: most studies indicate that clathrin is essential for endocytosis at hippocampal synapses³⁶, whereas ultrafast endocytosis is clathrin-independent. Here we demonstrate that even the ultrafast pathway requires clathrin to regenerate synaptic vesicles, not at the plasma membrane but rather at the synaptic endosome. It is probable that at some synapses and under some physiological conditions, clathrin acts at the plasma membrane to regenerate synaptic vesicles. Whether ultrafast endocytosis is a general mechanism or a specialization for synapses with high turnover will require further study.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 February; accepted 3 September 2014.

Published online 8 October 2014.

1. Heuser, J. E. & Reese, T. S. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* **57**, 315–344 (1973).
2. Maycox, P. R., Link, E., Reetz, A., Morris, S. A. & Jahn, R. Clathrin-coated vesicles in nervous tissue are involved primarily in synaptic vesicle recycling. *J. Cell Biol.* **118**, 1379–1388 (1992).
3. Takei, K. *et al.* Generation of coated intermediates of clathrin-mediated endocytosis on protein-free liposomes. *Cell* **94**, 131–141 (1998).
4. Takei, K., Mundigl, O., Daniell, L. & De Camilli, P. The synaptic vesicle cycle: a single vesicle budding step involving clathrin and dynamin. *J. Cell Biol.* **133**, 1237–1250 (1996).
5. Li, C. *et al.* Ca²⁺-dependent and -independent activities of neural and non-neural synaptotagmins. *Nature* **375**, 594–599 (1995).
6. Jorgensen, E. M. *et al.* Defective recycling of synaptic vesicles in synaptotagmin mutants of *Caenorhabditis elegans*. *Nature* **378**, 196–199 (1995).
7. Zhang, J. Z., Davletov, B. A., Südhof, T. C. & Anderson, R. G. W. Synaptotagmin I is a high affinity receptor for clathrin AP-2: implications for membrane recycling. *Cell* **78**, 751–760 (1994).

8. Nonet, M. L. *et al.* UNC-11, a *Caenorhabditis elegans* AP180 homologue, regulates the size and protein composition of synaptic vesicles. *Mol. Biol. Cell* **10**, 2343–2360 (1999).
9. Zhang, B. *et al.* Synaptic vesicle size and number are regulated by a clathrin adaptor protein required for endocytosis. *Neuron* **21**, 1465–1475 (1998).
10. Watanabe, S. *et al.* Ultrafast endocytosis at mouse hippocampal synapses. *Nature* **504**, 242–247 (2013).
11. Watanabe, S. *et al.* Ultrafast endocytosis at *Caenorhabditis elegans* neuromuscular junctions. *eLife* **2**, e00723 (2013).
12. Christie, J. M. & Jahr, C. E. Multivesicular release at Schaffer collateral-CA1 hippocampal synapses. *J. Neurosci.* **26**, 210–216 (2006).
13. Berndt, A. *et al.* High-efficiency channelrhodopsins for fast neuronal stimulation at low light levels. *Proc. Natl Acad. Sci. USA* **108**, 7595–7600 (2011).
14. Rizzoli, S. O. *et al.* Evidence for early endosome-like fusion of recently endocytosed synaptic vesicles. *Traffic* **7**, 1163–1176 (2006).
15. Royle, S. J., Bright, N. A. & Lagnado, L. Clathrin is required for the function of the mitotic spindle. *Nature* **434**, 1152–1157 (2005).
16. Lee, D., Lin, B.-J. & Lee, A. K. Hippocampal place fields emerge upon single-cell manipulation of excitability during behavior. *Science* **337**, 849–853 (2012).
17. Sik, A., Penttonen, M., Ylinen, A. & Buzsáki, G. Hippocampal CA1 interneurons: an *in vivo* intracellular labeling study. *J. Neurosci.* **15**, 6651–6665 (1995).
18. Macia, E. *et al.* Dynasore, a cell-permeable inhibitor of dynamin. *Dev. Cell* **10**, 839–850 (2006).
19. Park, R. J. *et al.* Dynamin triple knockout cells reveal off target effects of commonly used dynamin inhibitors. *J. Cell Sci.* **126**, 5305–5312 (2013).
20. Kane, R. E. Actin polymerization and interaction with other proteins in temperature-induced gelation of sea urchin egg extracts. *J. Cell Biol.* **71**, 704–714 (1976).
21. Jensen, V., Walaas, S. I., Hilfinger, S., Ruiz, A. & Hvalby, Ø. A delayed response enhancement during hippocampal presynaptic plasticity in mice. *J. Physiol. (Lond.)* **583**, 129–143 (2007).
22. Hoopmann, P. *et al.* Endosomal sorting of readily releasable synaptic vesicles. *Proc. Natl Acad. Sci. USA* **107**, 19055–19060 (2010).
23. Schikorski, T. Readily releasable vesicles recycle at the active zone of hippocampal synapses. *Proc. Natl Acad. Sci. USA* **111**, 5415–5420 (2014).
24. Miller, T. M. & Heuser, J. E. Endocytosis of synaptic vesicle membrane at the frog neuromuscular junction. *J. Cell Biol.* **98**, 685–698 (1984).
25. Heuser, J. & Reese, T. in *Fourth Intensive Study Program in the Neurosciences* 573–600 (M.I.T. Press, 1979).
26. Gisselsson, L. L., Matus, A. & Wieloch, T. Actin redistribution underlies the sparing effect of mild hypothermia on dendritic spine morphology after *in vitro* ischemia. *J. Cereb. Blood Flow Metab.* **25**, 1346–1355 (2005).
27. Hartmann-Petersen, R., Walmod, P. S., Berezin, A., Berezin, V. & Bock, E. Individual cell motility studied by time-lapse video recording: influence of experimental conditions. *Cytometry* **40**, 260–270 (2000).
28. Pyott, S. J. & Rosenmund, C. The effects of temperature on vesicular supply and release in autaptic cultures of rat and mouse hippocampal neurons. *J. Physiol. (Lond.)* **539**, 523–535 (2002).
29. Kushmerick, C., Renden, R. & von Gersdorff, H. Physiological temperatures reduce the rate of vesicle pool depletion and short-term depression via an acceleration of vesicle recruitment. *J. Neurosci.* **26**, 1366–1377 (2006).
30. Teng, H., Cole, J. C., Roberts, R. L. & Wilkinson, R. S. Endocytic active zones: hot spots for endocytosis in vertebrate neuromuscular terminals. *J. Neurosci.* **19**, 4855–4866 (1999).
31. Teng, H., Lin, M. Y. & Wilkinson, R. S. Macroendocytosis and endosome processing in snake motor boutons. *J. Physiol. (Lond.)* **582**, 243–262 (2007).
32. Micheva, K. D. & Smith, S. J. Strong effects of subphysiological temperature on the function and plasticity of mammalian presynaptic terminals. *J. Neurosci.* **25**, 7481–7488 (2005).
33. Renden, R. & von Gersdorff, H. Synaptic vesicle endocytosis at a CNS nerve terminal: faster kinetics at physiological temperatures and increased endocytotic capacity during maturation. *J. Neurophysiol.* **98**, 3349–3359 (2007).
34. von Gersdorff, H. & Matthews, G. Dynamics of synaptic vesicle fusion and membrane retrieval in synaptic terminals. *Nature* **367**, 735–739 (1994).
35. Neves, G., Gomis, A. & Lagnado, L. Calcium influx selects the fast mode of endocytosis in the synaptic terminal of retinal bipolar cells. *Proc. Natl Acad. Sci. USA* **98**, 15282–15287 (2001).
36. Granseth, B., Odermatt, B., Royle, S. J. & Lagnado, L. Clathrin-mediated endocytosis is the dominant mechanism of vesicle retrieval at hippocampal synapses. *Neuron* **51**, 773–786 (2006).
37. Balaji, J. & Ryan, T. A. Single-vesicle imaging reveals that synaptic vesicle exocytosis and endocytosis are coupled by a single stochastic mode. *Proc. Natl Acad. Sci. USA* **104**, 20576–20581 (2007).
38. Armbruster, M., Messa, M., Ferguson, S. M., De Camilli, P. & Ryan, T. A. Dynamin phosphorylation controls optimization of endocytosis for brief action potential bursts. *eLife* **2**, e00845 (2013).
39. Gad, H., Löw, P., Zotova, E., Brodin, L. & Shupliakov, O. Dissociation between Ca^{2+} -triggered synaptic vesicle exocytosis and clathrin-mediated endocytosis at a central synapse. *Neuron* **21**, 607–616 (1998).
40. Wenk, M. R. & De Camilli, P. Protein–lipid interactions and phosphoinositide metabolism in membrane traffic: insights from vesicle recycling in nerve terminals. *Proc. Natl Acad. Sci. USA* **101**, 8262–8269 (2004).
41. Takamori, S. *et al.* Molecular anatomy of a trafficking organelle. *Cell* **127**, 831–846 (2006).
42. Wucherpfennig, T., Wilsch-Bräuninger, M. & González-Gaitán, M. Role of *Drosophila* Rab5 during endosomal trafficking at the synapse and evoked neurotransmitter release. *J. Cell Biol.* **161**, 609–624 (2003).
43. Uytterhoeven, V., Kuenen, S., Kasprowitz, J., Miskiewicz, K. & Verstreken, P. Loss of skywalker reveals synaptic endosomes as sorting stations for synaptic vesicle proteins. *Cell* **145**, 117–132 (2011).
44. Gu, M. *et al.* AP2 hemicomplexes contribute independently to synaptic vesicle endocytosis. *eLife* **2**, e00190 (2013).
45. Heerssen, H., Fetter, R. D. & Davis, G. W. Clathrin dependence of synaptic-vesicle formation at the *Drosophila* neuromuscular junction. *Curr. Biol.* **18**, 401–409 (2008).
46. Kasprowitz, J. *et al.* Inactivation of clathrin heavy chain inhibits synaptic recycling but allows bulk membrane uptake. *J. Cell Biol.* **182**, 1007–1016 (2008).
47. Kononenko, N. L. *et al.* Clathrin/AP-2 mediate synaptic vesicle reformation from endosome-like vacuoles but are not essential for membrane retrieval at central synapses. *Neuron* **82**, 981–988 (2014).
48. Kittelmann, M. *et al.* *In vivo* synaptic recovery following optogenetic hyperstimulation. *Proc. Natl Acad. Sci. USA* **110**, E3007–E3016 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank D. Lorenz and A. Muenster-Wandowski for providing access to electron microscopy, J. Iwasa for the original model figure, S. Jorgensen for image processing, and C. Garner for discussions. We thank EMBO for providing the travel funds (S.W.). The research was funded by the US National Institutes of Health (NS034307 EMJ), European Research Council grant (249939 SYNGLUT CR), and German Research Council grants (Neurocore EXC 257 EMJ+CR, SFB 665 + SFB 958 CR). E.M.J. is an Investigator of the Howard Hughes Medical Institute and is an Alexander von Humboldt Scholar.

Author Contributions S.W., C.R., and E.M.J. conceived and designed experiments. S.W. and B.S.-K. performed the freezing experiments. S.W. performed electron microscopy imaging and analysis. T.T. designed shRNA constructs. T.T. and B.S.-K. generated lentivirus and performed biochemistry. T.T. and B.S.-K. performed immunocytochemistry. T.T., M.C.-P., and B.R.R. performed electrophysiology. A.F. prepared cell cultures. M.W.D. designed the stimulation device. S.W., T.T., M.C.-P., B.R.R., C.R., and E.M.J. wrote the manuscript. C.R. and E.M.J. provided funding, experiments were performed at the Charité, Berlin.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.M.J. (jorgensen@biology.utah.edu) or C.R. (christian.rosenmund@charite.de).

METHODS

Cell culture and lentivirus infection. Animals were handled according to the rules of the Berlin authorities and the animal welfare committee of the Charité Berlin, Germany. Hippocampi were dissected from newborn C57/BL6-N mice and cultured at 13×10^3 cells per cm^2 on 6 mm sapphire disks for 'flash-and-freeze' electron microscopy experiments and 25 mm coverslips for all other experiments as previously described¹⁰. Astrocytes were seeded a week in advance to generate a feeder layer. For autaptic cultures, microislands of astrocytes were prepared as previously described²⁸ and neurons were plated at 300 cells per cm^2 . For biochemistry, 10×10^4 cells per cm^2 were cultured without the feeder layer.

Lentivirus was produced as described previously⁵⁰. The cells were infected with lentivirus expressing ChetaTC¹⁰ on DIV1 and clathrin or non-specific shRNA on DIV7. Infection of clathrin shRNA at DIV1 caused severe loss of neurons at DIV14, and thus, cells were only infected for 1 week with clathrin shRNA for all the experiments. **shRNA constructs.** A mouse CHC-1 specific siRNA target sequence (5'-GTTG GTGACCGTTGTTATG-3') was obtained using Genscript siRNA Target Finder (<https://www.genscript.com/ssl-bin/app/rnai>) and cloned as shRNA into a lentiviral shuttle vector under the control of a U6 promoter. For the scrambled shRNA control sequence we adapted the scrambled siRNA sequence from ref. 15 and cloned it as shRNA (5'-TTCGACCCCTACTTCGTGG-3') into a lentiviral shuttle vector. Both sequences were subject to a BLAST search to ensure that mouse CHC-1 shRNA was specific and that the scrambled shRNA did not match any sequence. To identify infected cells, the shuttle vector contained a human Synapsin-1 promoter, driving an expression of a nuclear-targeted red fluorescent protein (NLS-RFP).

Western blots and immunocytochemical staining. For detection of clathrin levels by western blots, protein lysates were obtained from astrocyte-free mass cultures of hippocampal neurons. Briefly, cells were lysed using 50 mM Tris/HCl (pH 7.9), 150 mM NaCl, 5 mM EDTA, 1% Triton-X-100, 1% Nonidet P-40, 1% sodium deoxycholate, and protease inhibitors (cOmplete protease inhibitor cocktail tablet, Roche Diagnostics GmbH, Mannheim, Germany). Protein concentration was determined by BCA assay. Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes. Membranes were then incubated with rabbit anti-CHC-1 (1:500, Abcam, ab21679) and mouse anti-tubulin III (1:2,000, Sigma-Aldrich, T8660) antibodies overnight at 4 °C. After incubation with corresponding horseradish peroxidase-conjugated goat secondary antibodies (1 h at room temperature) and ECL Plus Western Blotting Detection Reagents (GE Healthcare Biosciences), chemiluminescent was imaged using a Vilber Lourmat Fusion FX7 detection system. Ratiometric quantification of signal intensities was measured with the supplied BIO-1D software package of the Fusion FX system. The signals are linear with the amount of protein lysate loaded (data not shown), suggesting that we can detect the clathrin signals in the knockdown without saturating the signals from the control.

Immunocytochemical staining was performed on autaptic cultures. One week after infection with the shRNA constructs, cells were washed once in PBS and fixed with 4% paraformaldehyde (PFA) for 10 min. After washing in PBS, cells were permeabilized with 0.1% PBS-Tween20 for 10 min and blocked in 5% NGS for 30 min. Subsequently cells were incubated with rabbit anti-CHC-1 (1:1,000, abcam, ab21679) and guinea pig anti-synaptophysin (1:1,000, Synaptic System) antibodies overnight at 4 °C. Primary antibodies were labelled with anti-rabbit Alexa Fluor 647 and anti-guinea pig Alexa Fluor 488 (each 1:500, Jackson Immuno-research Laboratories) for 1 h at room temperature. After washing, coverslips were mounted with Mowiol 4-88 anti-fade medium (Polysciences Europe GmbH, Eppelheim, Germany). Transduced cells, visible by nuclear RFP expression, were imaged using an Olympus IX81 microscope equipped with a Princeton camera and controlled by Metamorph software, with same acquisition settings for all groups. Ratiometric quantification of clathrin signals over synaptophysin signals was conducted in the automated fashion using ImageJ (National Institute of Health, Bethesda, MD) with custom-written macros. In short, the program first subtracts uneven illumination using a 'rolling ball' function with the radius set at 30 pixels (http://fiji.sc/Rolling_Ball_Background_Subtraction). Then, the synapses are defined by thresholding the synaptophysin signals at 15%. The binary image of synaptophysin was used as the regions of interest. The average intensities of fluorescence signals from clathrin heavy chains and synaptophysin were measured from those locations and were divided to obtain ratio between those two proteins. A total of 28 images from the control shRNA cells and 30 images from the clathrin knockdown cells ($n = 2$ cultures) were used.

Electrophysiology. Whole cell patch-clamp and cell-attached voltage-clamp recordings were performed as previously described¹⁰. The extracellular solution contained 140 mM NaCl, 2.4 mM KCl, 10 mM HEPES, 10 mM Glucose (pH adjusted to 7.3 with NaOH, 300 mOsm). For a single light stimulus experiment, 4 mM CaCl_2 and 1 mM MgCl_2 were added to the solution, while 2 mM CaCl_2 and 1 mM MgCl_2 were added for high-frequency stimulation experiments. All recordings were performed at 34 °C.

For recordings in autapses, neurons at DIV15–17 were clamped at -70 mV with a Multiclamp 700B amplifier (Molecular Devices) under control of Clampex 9 (Molecular Devices). The extracellular solution contained 2 mM CaCl_2 and 4 mM MgCl_2 . The patch pipette solution contained 136 mM KCl, 17.8 mM HEPES, 1 mM EGTA, 0.6 mM MgCl_2 , 4 mM ATP-Mg, 0.3 mM GTP-Na, 12 mM phosphocreatine and 50 U per ml phosphocreatine kinase (300 mOsm, pH 7.4). EPSCs were evoked by a brief 2 ms somatic depolarization to 0 mV. Readily-releasable pool size was determined by measuring the charge transfer of the transient synaptic current induced by a pulsed 5 s application of hypertonic solution (500 mM sucrose in extracellular solution). Vesicular release probability was calculated as the ratio of the charge from the evoked EPSC (integrated for 1 s) and the readily-releasable pool size.

Electron microscopy. 'Flash-and-freeze' experiments were performed as previously described¹⁰. Briefly, sapphire disks with cultured cells were mounted in the freezing chamber of the high-pressure freezer (HPM100, Leica), which was set at 34 °C. The temperature 34 °C was chosen to match the temperature of electrophysiological recordings. To minimize the exposure to room temperature, a Petri dish containing the sapphire disk was placed on a 37 °C cryopack while mounting, and the transparent polycarbonate sample cartridges were stored between cryopacks equilibrated to 37 °C. Immediately after the sapphire disk was mounted on the sample holder, recording solution kept at 37 °C was applied to the specimen and the cartridge was inserted into the freezing chamber and allowed to equilibrate to 34 °C for 30 s. For experiments at 37 °C (Extended Data Fig. 2), a temperature-controlled chamber (Leica) was placed around the specimen loading table of the high-pressure freezer. After transferring in a Petri dish, the cells were allowed to recover in the 37 °C chamber for 2 min and then mounted onto the specimen carriers for freezing. The freezing chamber was set at 37 °C in these experiments. Using a home-built light stimulation controller, we applied either 1 or 10 light pulses (20 Hz) to the specimens. The controls for each experiment were always frozen on the same day from the same animal. Each light pulse was for 10 ms. Under these conditions we observe multivesicular release; we estimate that up to 4 synaptic vesicles are released per active zone (data not shown and ref. 10). We set the device so that the samples were frozen at 15, 30, 50, 100, 300, 1,000, 3,000, 10,000 or 20,000 ms after the initiation of the first stimulus¹⁰.

For ferritin-loading experiments, cationized ferritin (Sigma-Aldrich) was added in the recording solution at 0.25 mg ml^{-1} . The calcium concentration was reduced to 1 mM to suppress spontaneous activity during the loading. The cells were incubated in the solution for 5 min at 37 °C. For Pitstop 2 experiments, 30 μM Pitstop 2 was also included in the solution, and the cells were only incubated for 2 min, as longer incubations sometimes yielded dead cells. After ferritin incubation, the cells were immersed in the recording solution containing either 4 mM Ca^{2+} for a single stimulus experiment or 2 mM Ca^{2+} for high-frequency stimulation. The change in calcium concentrations from 1 mM to 4 mM increases the rate of miniature EPSCs and thus may contribute to the background ferritin loading before the experiments. In our experiments, about 1% of synaptic profiles contained endosomes and synaptic vesicles that were ferritin-positive without stimulation. The absence of ferritin-positive coated pits on the plasma membrane in the unstimulated controls suggests that ferritin passes through endosomes even during spontaneous activity. 3 μM NBQX and 30 μM bicuculline were also included in the recording media to minimize the recurrent network activity¹⁰. Ferritin-positive synaptic vesicles were only found in 5–7% synaptic profiles scored before stimulation using this protocol.

Following high-pressure freezing, samples were transferred into a vial containing 1% osmium tetroxide (EMS), 1% glutaraldehyde (EMS), 1% milliQ water, in anhydrous acetone (EMS). The freeze-substitution was performed in AFS2 (Leica) with the following program: -90 °C for 5–7 h, 5 °C per hour to -20 °C, 12 h at -20 °C, and 10 °C per hour to 20 °C. Following en bloc staining with 0.1% uranyl acetate, the samples were infiltrated and embedded into epon and cured for 48 h in a 60 °C oven. Serial 40-nm sections were cut using a microtome (Leica UCT) and collected onto formvar-coated single-slot grids. Sections were stained with 2.5% uranyl acetate before imaging. For ferritin experiments, sections were not stained after sectioning to improve contrast of ferritin molecules in our images—this, in turn, might have compromised our ability to distinguish clathrin-coated vesicles. Approximately 150 synaptic profiles were collected from a single section from each specimen, and the experiments were repeated with second cultures in each case (for detailed n values, see Supplementary Table 1). The sample size was chosen based on the previous experiments that allowed us to acquire a sufficiently large set of data for statistical analysis¹⁰. The total numbers of synaptic profiles and tomograms analysed for these experiments are 10,514 and 361, respectively. Note that there is some variability from experiment to experiment; this is probably due to the short lifetimes of these structures. The synaptic profiles were chosen randomly to sample unbiased populations. Active zones were defined as regions juxtaposed to a post-synaptic density. Docked vesicles are defined as those directly in contact with membrane. The large endocytic vesicles are defined as vesicles larger than ~ 50 nm by visual inspection and within 50 nm of active zone, measured in ImageJ. Endosomes

are defined as membrane-bound organelles that are in the centre of the bouton and larger than 100 nm by visual inspection. Vesicular compartments with coated buds were also categorized as endosomes, which appeared frequently in our 3 s time points. Vesicles are scored as clathrin-coated only if distinctive coats were visible which can lead to underscoring (Fig. 1c and Extended Data Fig. 2c). Furthermore, the chance of capturing a coated vesicle in a given synaptic profile is low compared to an endosome due to its size. These factors will lead to an underestimate of the number of coated vesicles observed in profiles. The morphometry was performed blind using custom-written ImageJ macro and Matlab scripts (Watanabe, Davis, and Jorgensen).

For electron tomography, 100–200 nm thick sections were collected on pioloform-coated single-slot grids. Sections were post-stained with uranyl acetate as described for thin sections. 10-nm gold particles were sparsely applied to both sides of the grids by incubating the grids in drops of solution containing 5.7×10^{12} particles per ml (<http://microspheres-nanospheres.com/>) for 2 min. 17–50 tilt-series ($\pm 65^\circ$) were collected from each sample using FEI TF20 electron microscope. For serial tomograms, low magnification images acquired from each section on a grid were used to locate the same synapses in the serial sections. Typically, the entire synapses span 4 to 5 sections, when cut at 200 nm. The tomograms were generated from tilt-series using IMOD. The tomograms were segmented and reconstructed using Amira or TrakEM2.

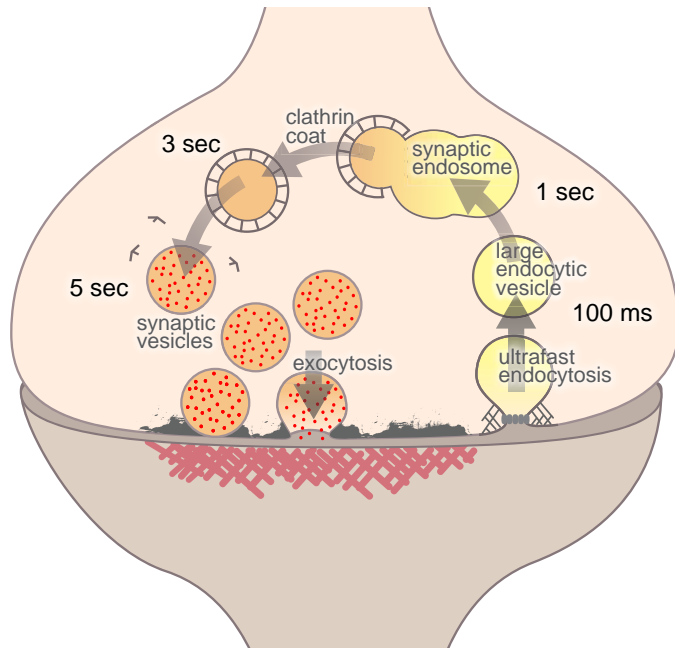
Membrane calculations. The amount of membrane on large endocytic vesicles and endosomes was calculated to determine the number of synaptic vesicle equivalents contained in these structures. These structures were scored blind and were distinguished by position in the bouton, by size, and by shape. For large endocytic vesicles in single stimulus experiments (Fig. 1), the mean diameter was 80.6 ± 1.5 nm. The surface area of the large endocytic vesicles is therefore about $20,000 \text{ nm}^2$, as the surface area of a synaptic vesicle is $5,300 \text{ nm}^2$, large endocytic vesicles correspond to 4 synaptic vesicle equivalents. In the RNA interference experiments after a single stimulus, the size of large endocytic vesicles was 82.6 ± 4.5 nm (4 synaptic vesicle

equivalents) in the control and 74.4 ± 2.0 nm (3 synaptic vesicle equivalents) in the clathrin knockdown. In the RNA interference experiments after high-frequency stimulation, the size of large endocytic vesicles was 85.5 ± 2.2 nm (4 synaptic vesicle equivalents) in the control and 70.1 ± 1.4 nm (3 synaptic vesicle equivalents) in the knockdown.

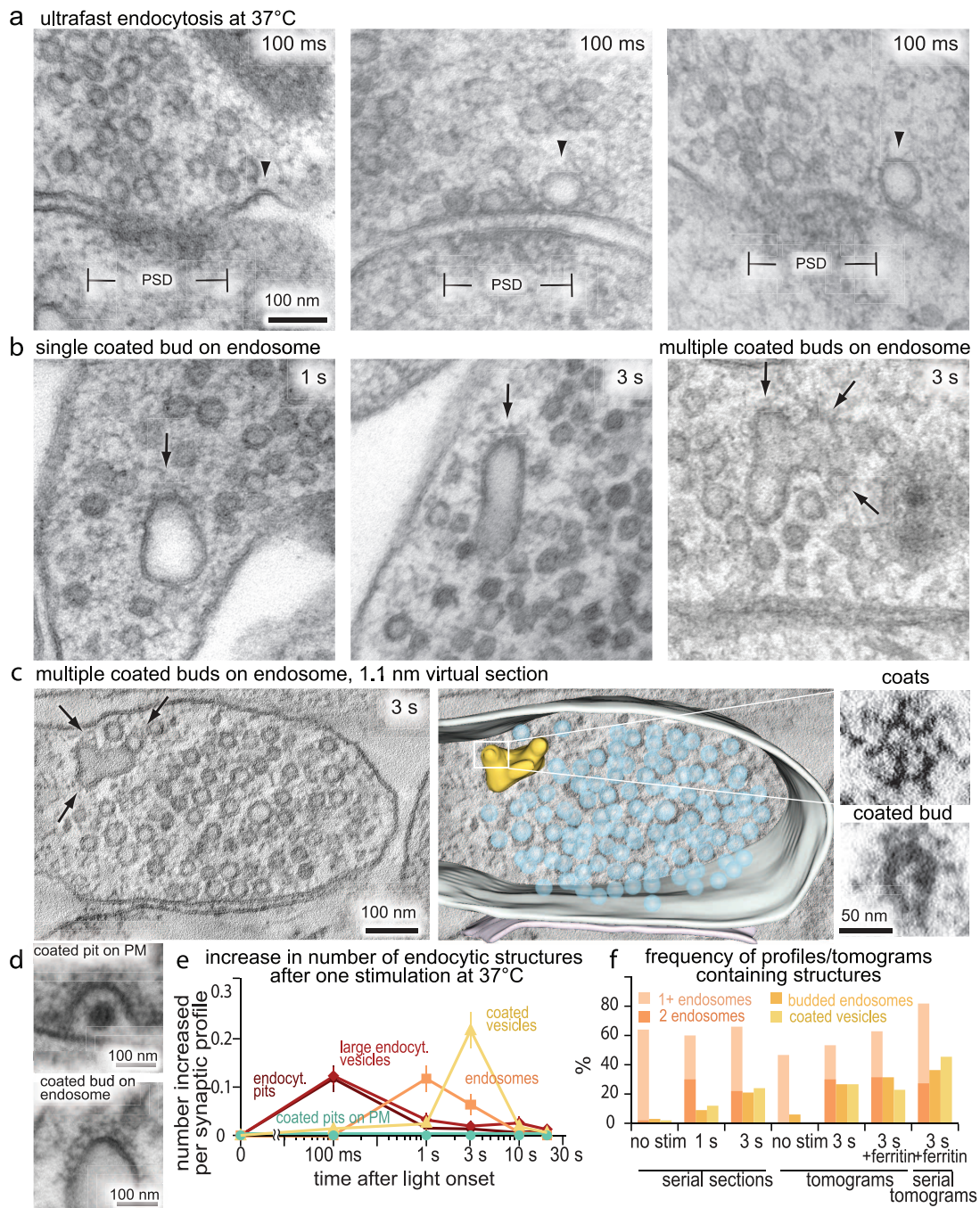
The surface area of endosomes was calculated by measuring the diameter and assuming a spherical shape. This calculation will lead to an underestimate of the surface area of endosomes because the membrane in budding endosomes is convoluted. To minimize this error, endosomes with coated buds were not included in the calculations. After a single stimulus (Fig. 1), the mean diameter of endosomes was 116.4 ± 3.0 nm (8 synaptic vesicle equivalents). In the ferritin experiments with a single stimulus, the diameter of endosomes was 108.7 ± 4.2 nm (7 synaptic vesicle equivalents) in the control and 93.1 ± 2.8 nm (5 synaptic vesicle equivalents) in the clathrin knockdown. In the ferritin experiments with multiple stimuli, the diameter of endosomes was 111.8 ± 4.1 nm (8 synaptic vesicle equivalents) in the control and 100.9 ± 2.6 nm (6.5 synaptic vesicle equivalents) in the clathrin knockdown. The smaller size of endocytic structures in the clathrin knockdowns is probably due to the reduced neurotransmission in these cells. To confirm the estimates of surface area, we measured the surface area of complete endosomes in tomograms of wild-type synapses. Most endosomes were spherical with the diameter of 111.2 ± 5.3 nm ($39,000 \text{ nm}^2$ or 7.3 synaptic vesicle equivalents). The surface area of budded endosomes was measured to be $\sim 41,000 \text{ nm}^2$ or 7.7 synaptic vesicle equivalents.

Statistics. For detailed numbers and statistical analysis, see Supplementary Table 1.

49. von Kleist, L. *et al.* Role of the clathrin terminal domain in regulating coated pit dynamics revealed by small molecule inhibition. *Cell* **146**, 471–484 (2011).
50. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868–872 (2002).



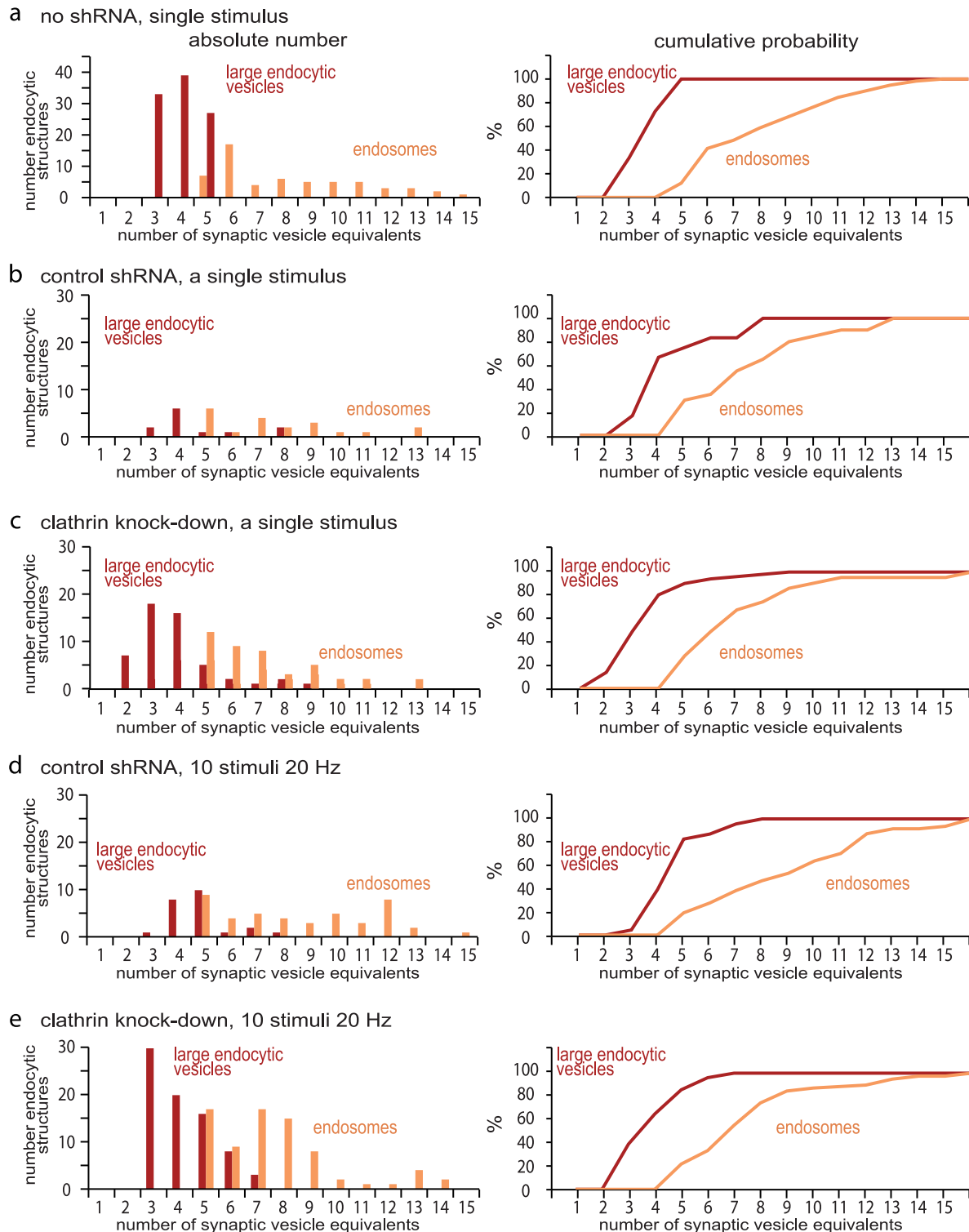
Extended Data Figure 1 | Ultrafast endocytosis regenerates synaptic vesicles in a two-step process. Ultrafast endocytosis removes membrane added by vesicle fusion at the lateral edge of the active zone. Large endocytic vesicles then fuse to endosomes. Endosomes are resolved into synaptic vesicles via a clathrin-dependent process. Newly formed synaptic vesicles can be recruited back to the active zone.



Extended Data Figure 2 | Synaptic vesicles are regenerated from endosomes at 37 °C.

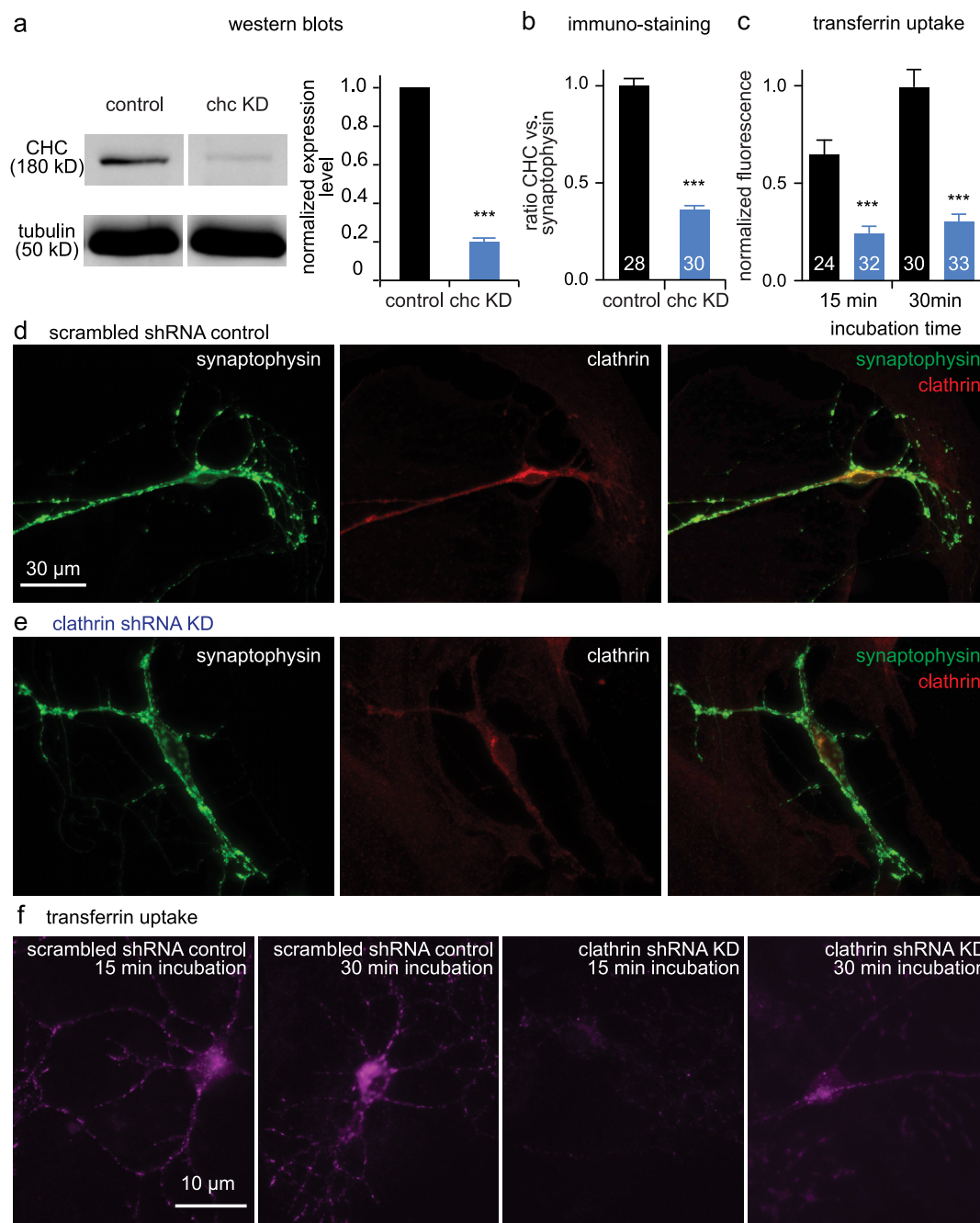
Hippocampal synapses were stimulated once and frozen at the indicated times. The experiments were performed at 37 °C in the presence of 4 mM external Ca^{2+} . **a**, Electron micrographs showing invaginations and large endocytic vesicles (arrowheads) recovered via ultrafast endocytosis. **b**, Micrographs showing single coated buds (left, middle) or multiple buds (right) forming on an endosome. Arrows indicate coated buds on an endosome. **c**, Virtual section from an electron tomogram (left) and reconstruction (middle) showing a synaptic endosome with buds following a single stimulus. Arrows indicate coated buds on an endosome. The average intensity of coated buds from the top 20 nm (right, top) and the bottom 40 nm (right, bottom) is shown. Clathrin-cages can be preserved in our fixation and are visible in the tomogram. We found a total of 32 endosomes in these reconstructions (14 endosomes in the unstimulated control and 28 endosomes 3 s after stimulation). Of the total 32 endosomes, none were connected to the plasma membrane or showed evidence of a truncated tubule extending from the endosomal membrane. Of the 14 unstimulated endosomes, 8 were contained within single tomograms, and are therefore unambiguously closed

on both ends. Of the 28 endosomes in stimulated synapses, 16 endosomes were contained within single tomograms so that it was clear that no attachment to the plasma membrane was possible. **d**, Examples of a coated pit on the plasma membrane (top) and a coated bud on an endosome (bottom). Note that the morphology of the coats is similar. **e**, Increase in the number of each endocytic structure per synaptic profile after a single stimulus at 37 °C. The prevalence of large endocytic vesicles and endosomes is followed by an increase in the number of clathrin-coated vesicles. Coated pits were not observed on the plasma membrane (PM). **f**, Frequency of profiles or tomograms that contain endosomal structures at 37 °C. Approximately 60% of unstimulated synapses contained one endosome. One second after stimulation, 60% of the synapses contained at least one endosome, and half of those synapses contained two endosomes. Three seconds after stimulation, ~30% of the synapses contained budded endosomes and clathrin-coated vesicles, suggesting that those synapses were active. The standard error of the mean is shown in each graph. At least 146 synaptic profiles were analysed from each time point. For n values, detailed numbers and statistical analysis, see Supplementary Table 1.



Extended Data Figure 3 | Large endocytic vesicles likely fuse to become synaptic endosomes. **a–e**, Histograms (left) and cumulative plots (right), showing the size of large endocytic vesicles (red) and endosomes (orange) after one stimulus from control cells without ferritin (**a**), scrambled shRNA infected cells (**b**), and clathrin knockdown cells (**c**) both with ferritin. Ten stimuli were applied to scrambled shRNA (**d**) or clathrin knockdown cells (**e**). The large endocytic vesicles are defined as those that are within 50 nm of active zone and larger than a synaptic vesicle by visual inspection. Endosomes are defined as large vesicles greater than 50 nm from the active zone (often in the centre of the bouton) and are larger than ~100 nm. Any vesicular compartment that has

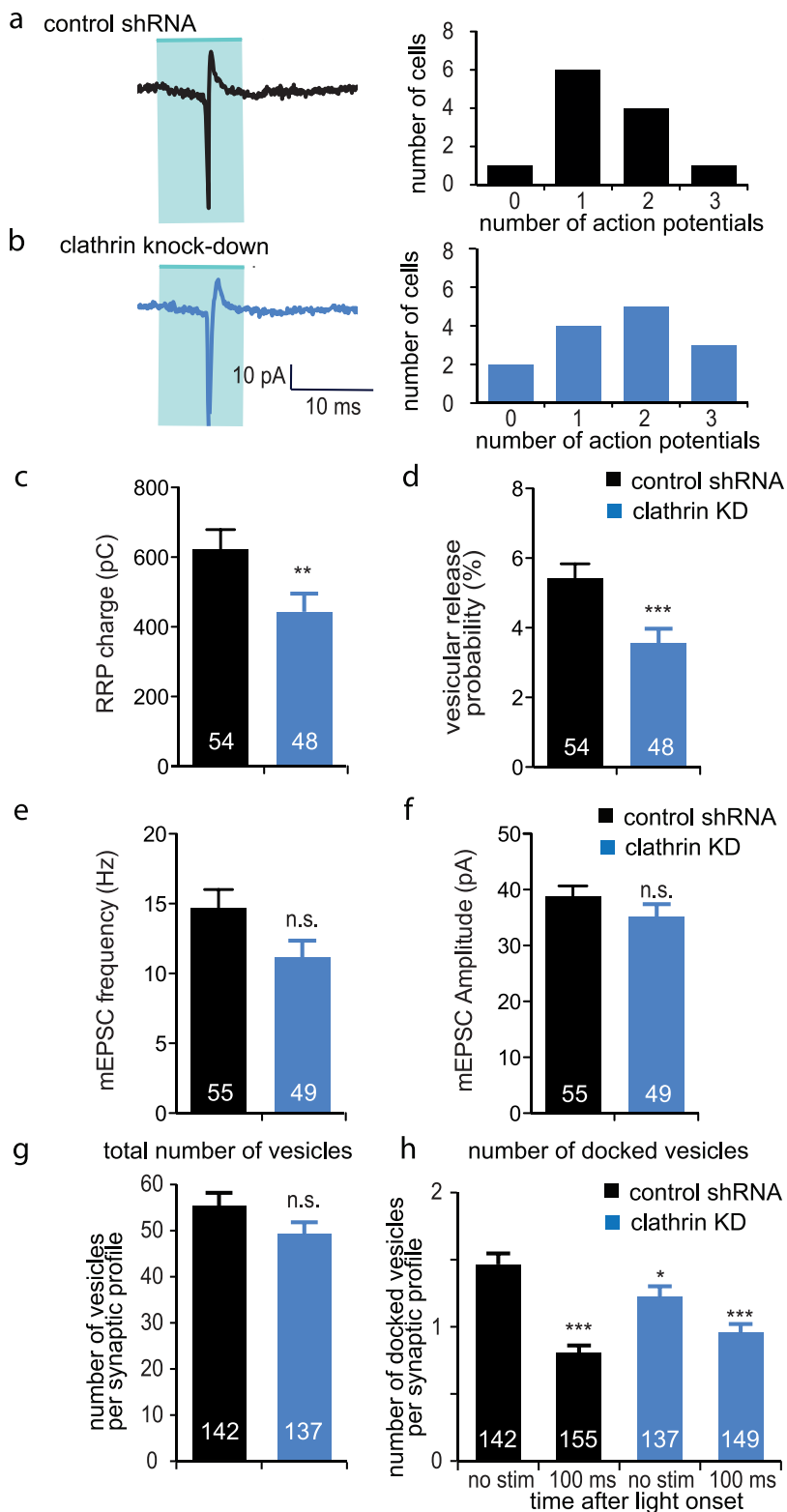
coated buds in the centre of the bouton is also categorized as an endosome. The numbers of endocytic structures in (**a**) represent all the structures scored from 100 ms and 300 ms time points. The number of large vesicles and endosomes in **b–e** represent all the ferritin-positive structures from later time points (3, 10, and 20 s). In the control shRNA (**b**), the number of large endocytic vesicles and endosomes has declined by these late time points, whereas the ferritin is trapped in large endocytic vesicles and synaptic endosomes in the clathrin knockdown experiments. Because ferritin passes from large endocytic vesicles to even larger budded endosomes, it is probable that the endocytic vesicles fuse with either each other or an existing compartment.



Extended Data Figure 4 | Clathrin shRNA reduces clathrin expression.

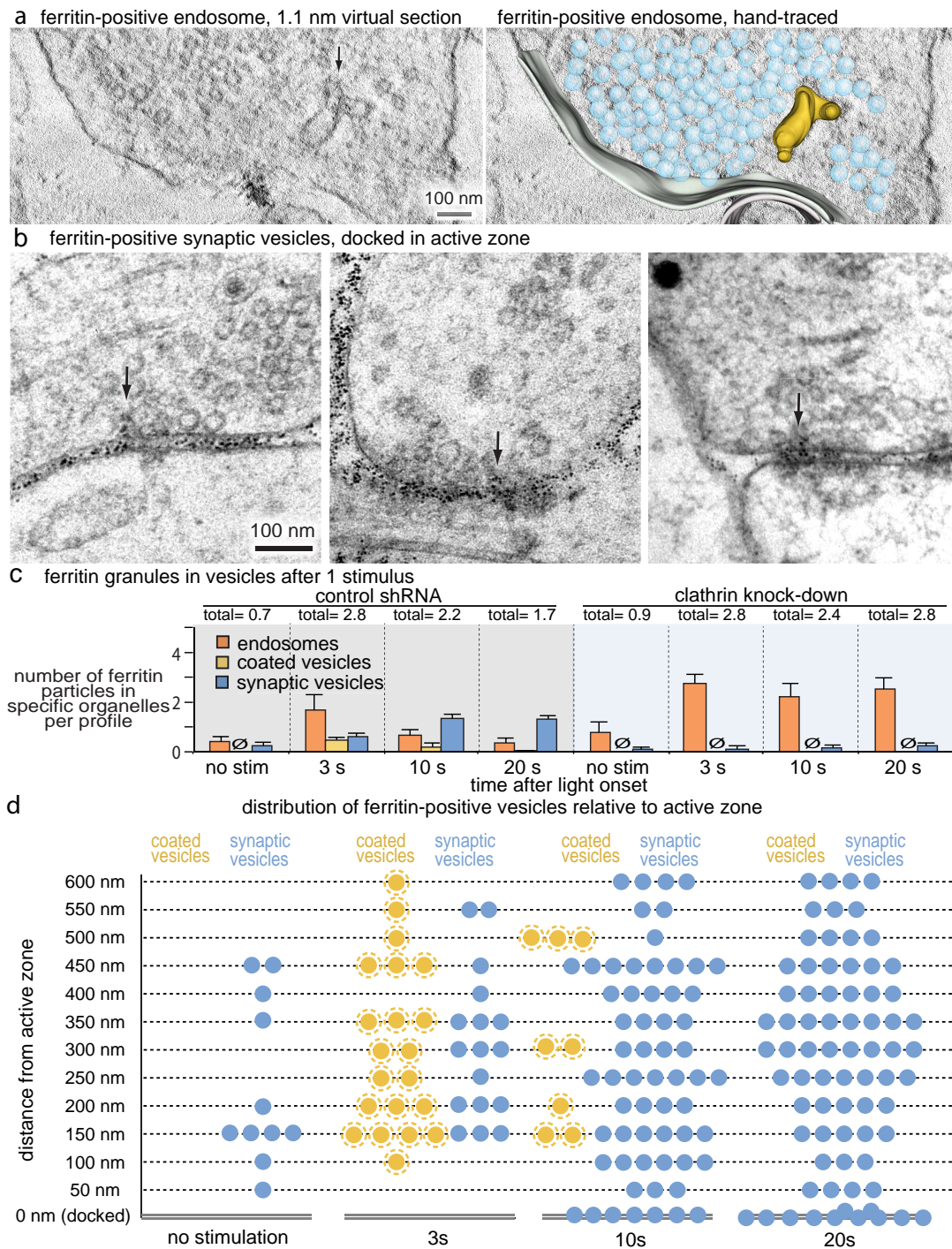
a, Left, western blot showing clathrin levels after one-week expression of a scrambled shRNA control or clathrin heavy chain shRNA in cultured hippocampal neurons. Right, an 80% reduction was observed ($n = 3$, $P < 0.001$, paired t -test). **b**, Normalized ratio of clathrin heavy chain and synaptophysin fluorescence in control and clathrin knockdown (chc KD) cultures. The clathrin/synaptophysin ratio is reduced to 64% in the knockdown cells. **c**, The mean fluorescence intensity (normalized to 30 min) representing the amount of

transferrin uptake in control and knockdown cells. Transferrin uptake was reduced by 66% in the knockdown cells. **d**, **e**, Fluorescence images of immunocytochemical staining of hippocampal autaptic cultures using anti-synaptophysin (left), anti-clathrin heavy chain (middle), and merge in control (**d**) and clathrin knockdown cultures (**e**). **f**, Example micrographs of hippocampal mass cultures showing transferrin uptake. The standard error of the mean is shown. *** $P < 0.001$. For detailed numbers and statistical analysis, see Supplementary Table 1.



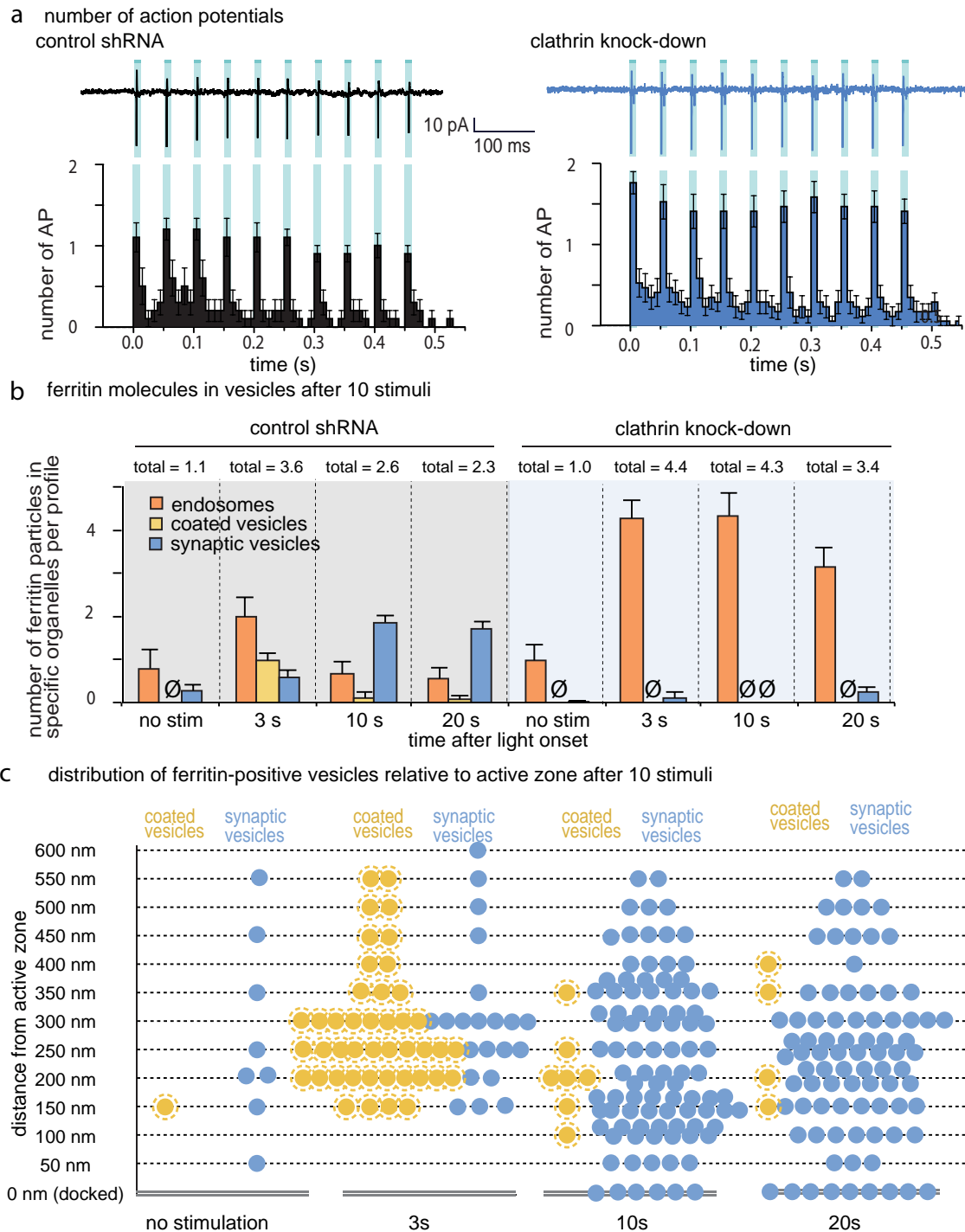
Extended Data Figure 5 | Exocytic machinery is intact in the clathrin knockdown cells. **a, b**, Sample traces from cell-attached voltage clamp during light stimulation in control (**a**) and clathrin knockdown (**b**). Number of action potentials triggered during the 10 ms light pulse is shown to the right. **c**, Readily-releasable pool (RRP) in control and clathrin knockdown cells, defined by brief application of 500 mM sucrose to autaptic neurons (control: 622 ± 56 pC, knockdown: 443 ± 52 pC, $P < 0.01$). **d**, Vesicular release probability (Pvr) in these cells (control $5.4 \pm 0.4\%$, $n = 54$; knockdown $3.5 \pm 0.4\%$, $n = 48$; $P < 0.001$). **e, f**, Average miniature EPSC (mEPSC) frequency (**e**) and amplitude (**f**). No change was observed in knockdown cells.

g. Average number of synaptic vesicles per synaptic profile ($n = 142$ synapses for control and 137 for knockdown). **h.** Average number of docked vesicles in active zones per synaptic profile (control: no stimulation, 1.5 ± 0.1 , $n = 142$ synapses; 100 ms, 0.8 ± 0.1 , $n = 142$ synapses; knockdown: no stimulation, 1.2 ± 0.1 , $n = 137$ synapses; 100 ms after stimulation, 1.0 ± 0.1 , $n = 149$ synapses). The fraction of docked vesicles that fuse is greatly reduced by clathrin knockdown. *P* values are calculated against the unstimulated control shRNA cells. The standard error of the mean is shown in each graph. *** $P < 0.001$, ** $P < 0.01$, and * $P < 0.05$, respectively. n.s., not significant.



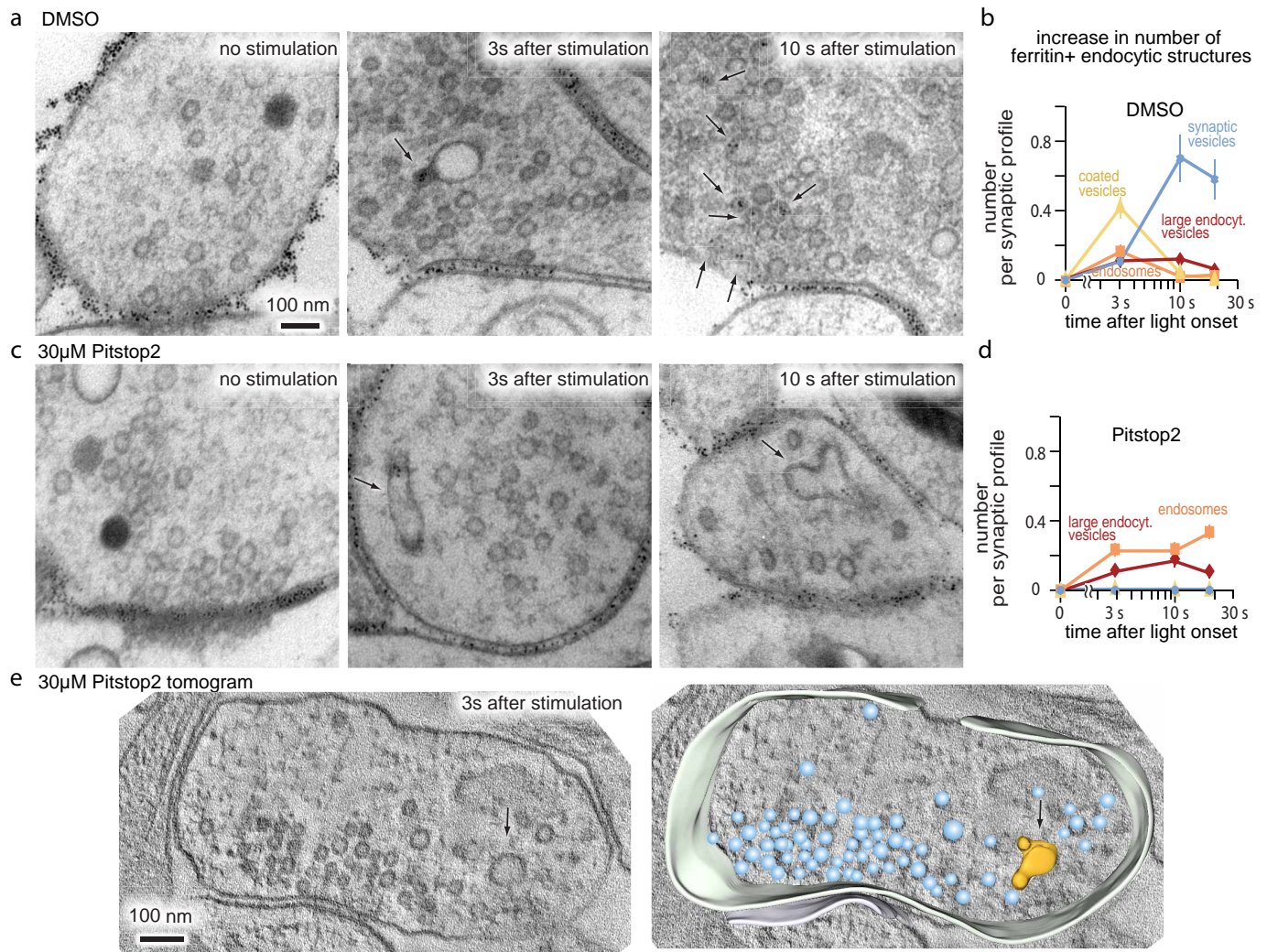
Extended Data Figure 6 | Following a single stimulus, clathrin is required at endosomes to regenerate synaptic vesicles. **a**, Virtual section from an electron tomogram (left) and a reconstruction (right) showing a budded synaptic endosome containing ferritin particles in the scrambled shRNA control cell. We found a total of 33 endosomes in these reconstructions. Of these 33 endosomes, none were connected to the plasma membrane or showed evidence of a tubule extending from the endosomal membrane. A total of 17 of these 33 total endosomes were fully contained within the 200 nm tomogram. Of these 33 endosomes, 16 were ferritin-positive, and 8 of these 16 endosomes were fully contained in the tomogram. **b**, Micrographs showing ferritin-positive synaptic vesicles docked to active zone 10–20 s after stimulation. **c**, Average number of ferritin particles in large endocytic vesicles, clathrin-coated vesicles, and synaptic vesicles per synaptic profile examined. At least 134 synapses were analysed per time point. Ferritin progresses to synaptic vesicles in the control, but is trapped in large endocytic vesicles or endosomes in the clathrin

knockdown. The fraction of synaptic profiles with ferritin was 27% for the control and 31% in the knockdown, suggesting that 70% of the synapses were silent. The mean number of ferritin particles found in an individual endosome, clathrin-coated vesicle, and synaptic vesicle, are 9.3 ± 1.0 , 2.0 ± 0.2 , and 1.9 ± 0.2 , respectively. The total number of ferritin particles (indicated above), declined by 40% in the control relative to the 1 s time point but not in the knockdown, either due to the fusion of the newly formed synaptic vesicles or by return of excess membrane to the surface of the synapse. The standard error of the mean is shown. **d**, Distribution of ferritin-positive clathrin-coated vesicles (yellow) and synaptic vesicles (blue) relative to the active zone at defined time points after stimulation in the control cells. Numbers are binned by 50 nm. The first bin '0 nm' means vesicles are docked in active zone. Endosomes are found at 285 ± 38 nm from the active zone. Note that the data in this figure represent further analysis of the data shown in Fig. 3.



Extended Data Figure 7 | Following high-frequency stimulation, clathrin is required at endosomes to regenerate synaptic vesicles. **a**, Average number of action potentials in 10 ms bins relative to light pulses during high-frequency stimulation (10 stimuli at 20 Hz, 0.5 s). Sample traces are shown above. Each light pulse triggered at least one action potential in both control and clathrin shRNA-treated cultures. **b**, Average number of ferritin molecules in large endocytic vesicles, clathrin-coated vesicles, and synaptic vesicles per profile examined. Ferritin is transferred from large endocytic vesicles to synaptic vesicles in the control but is trapped in large endocytic vesicles or endosomes in the clathrin knockdown. The number of profiles with ferritin particles after stimulation was 34% in the control and 36% in the clathrin knockdown, suggesting that 65% of the synapses were silent. On average, the number of

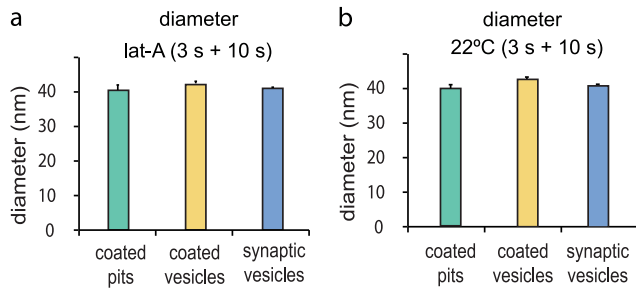
ferritin molecules found in an individual endosome, clathrin-coated vesicle, and synaptic vesicle, are 9.2 ± 1.0 , 1.9 ± 0.3 , and 2.0 ± 0.2 . At least 142 synapses were analysed per time point. The total number of ferritin particles (indicated above each time point), declined by 36% in the control, and by 23% in the knockdown relative to the 1 s time point. **c**, Distribution of ferritin-positive clathrin-coated vesicles (yellow) and synaptic vesicles (blue) relative to the active zone at defined time points after stimulation in the control shRNA cells. Numbers are binned by 50 nm. The first bin '0 nm' means vesicles are docked at the active zone. Endosomes are found at 286 ± 43 nm from the active zone. The standard error of the mean is shown in each graph. Note that this figure represents further analysis of the data from Fig. 4.



Extended Data Figure 8 | Pitstop 2 blocks regeneration of synaptic vesicles from endosomes after high-frequency stimulation. Pitstop 2 is an inhibitor of clathrin terminal domain and blocks clathrin-mediated endocytosis⁴⁹.

a, c, Electron micrographs showing ferritin containing vesicles in DMSO-treated (**a**) and Pitstop-2-treated cells (**c**) at different time points after stimulation. Ferritin is found in large vesicles after stimulation (middle) and in synaptic vesicles (right) in control, but it is trapped in endosomes in the Pitstop-2-treated cells. **b, d,** Average increase in ferritin-positive structures per synaptic profile in DMSO-treated (**b**) or Pitstop-2-treated cells (**d**). Ferritin progressed to synaptic vesicle-like structures in the control but remained in

endosomes or large endocytic vesicles in Pitstop-2-treated cells. Clathrin-coated pits on the plasma membrane were not present at any time point and were not plotted. At least 140 synapses were analysed per time point. **e,** Virtual section from an electron tomogram (left) and a reconstruction (right) showing a synaptic endosome with buds following 10 stimuli at 20 Hz. Of 32 tomograms reconstructed from the 3 s time point, 25 of them showed at least one endosome in the terminal, and 7 showed budded endosomes. None of these endosomes were connected to the plasma membrane. The standard error of the mean is shown in each graph. For detailed numbers and statistical analysis, see Supplementary Table 1.



Extended Data Figure 9 | Synaptic vesicles are regenerated directly from the plasma membrane in the absence of ultrafast endocytosis. a, b, Average diameter of clathrin-coated pits, clathrin-coated vesicles, and synaptic vesicles in the latrunculin-A treated cells (**a**) or cells incubated at 22 °C for 5 min (**b**). The diameter of these structures is similar suggesting a precursor-product relationship. Diameter of coated pits was determined by the full-width at the half maximum depth of the pit. For detailed numbers and statistical analysis, see Supplementary Table 1.

The expanding fireball of Nova Delphini 2013

G. H. Schaefer¹, T. ten Brummelaar¹, D. R. Gies², C. D. Farrington¹, B. Kloppenborg², O. Chesneau³, J. D. Monnier⁴, S. T. Ridgway⁵, N. Scott¹, I. Tallon-Bosc⁶, H. A. McAlister², T. Boyajian⁷, V. Maistro⁸, D. Mourard³, A. Meilland³, N. Nardetto³, P. Stee³, J. Sturmman¹, N. Vargas¹, F. Baron², M. Ireland⁹, E. K. Baines¹⁰, X. Che⁴, J. Jones², N. D. Richardson¹¹, R. M. Roettenbacher⁴, L. Sturmman¹, N. H. Turner¹, P. Tuthill⁸, G. van Belle¹², K. von Braun¹³, R. T. Zavala¹⁴, D. P. K. Banerjee¹⁵, N. M. Ashok¹⁵, V. Joshi¹⁵, J. Becker^{4,16} & P. S. Muirhead¹⁷

A classical nova occurs when material accreting onto the surface of a white dwarf in a close binary system ignites in a thermonuclear runaway^{1,2}. Complex structures observed in the ejecta at late stages^{3–5} could result from interactions with the companion during the common-envelope phase^{6,7}. Alternatively, the explosion could be intrinsically bipolar, resulting from a localized ignition on the surface of the white dwarf⁸ or as a consequence of rotational distortion^{9,10}. Studying the structure of novae during the earliest phases is challenging because of the high spatial resolution needed to measure their small sizes¹¹. Here we report near-infrared interferometric measurements of the angular size of Nova Delphini 2013, starting one day after the explosion and continuing with extensive time coverage during the first 43 days. Changes in the apparent expansion rate can be explained by an explosion model consisting of an optically thick core surrounded by a diffuse envelope. The optical depth of the ejected material changes as it expands. We detect an ellipticity in the light distribution, suggesting a prolate or bipolar structure that develops as early as the second day. Combining the angular expansion rate with radial velocity measurements, we derive a geometric distance to the nova of 4.54 ± 0.59 kiloparsecs from the Sun.

Nova Delphini 2013 (also known as V339 Del) was discovered¹² by K. Itagaki on 14 August 2013 Universal Time (UT) at 14:01. We began an intensive observing campaign to measure the size of the nova with the CHARA Array¹³, an optical/infrared interferometer located on Mount Wilson, California. Our observations began within 15 h of the discovery and within 24 h of the detonation itself. We measured the size of the expanding ejecta as the nova rose to peak brightness and continued monitoring it for a total of 27 nights between UT 2013 August 15 and September 26.

We measured the angular diameter of Nova Del 2013 by fitting a uniformly bright circle to the visibility amplitudes of the interference fringes recorded during each night (Extended Data Table 1). We plot the expansion curve in Fig. 1. The dotted line shows a linear fit to the angular diameters during the first 27 days after the explosion. The inset panel shows an apparent deceleration during the first four nights compared with lines of constant velocity. During the last week, the measurements show a large jump in the effective size of the nova compared with an extrapolation of the linear fit. For comparison, we show the visible and infrared light curves in Fig. 2.

The apparent changes in slope of the expansion curve can be explained by a two-component model consisting of an optically thick pseudo-photosphere surrounded by an optically thin halo. We approximated

the intensity distribution projected on the sky using a uniformly bright, circular core surrounded by a circular ring. We fitted the two-component model to the interferometric visibilities and minimized the total χ^2 simultaneously across all nights. We fixed the time of detonation at modified Julian day $t_0 = \text{MJD } 56518.277$, computed by extrapolating the first two pre-discovery photometric measurements¹⁴ back to the quiescent flux^{15,16}. From our interferometric data, we measured an expansion rate of 0.156 ± 0.003 milli-arcseconds (mas) per day for the core diameter and a size ratio of 1.73 ± 0.02 between the outer ring and core. We allowed the flux ratio between the ring and core to vary on a nightly basis.

Figure 3 shows how the percentage of light from the ring changes over time. During the first two nights, the core and ring have similar surface

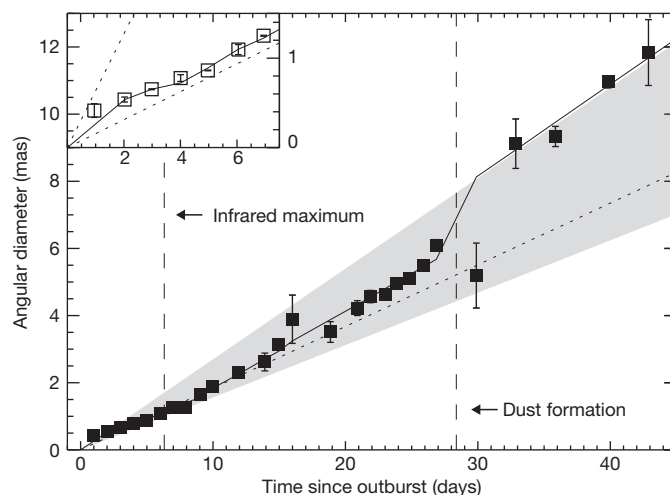


Figure 1 | Expansion curve of Nova Del 2013. The angular diameters were measured by fitting a circular disk to the interferometric data. The dotted line shows a linear fit for days 0–27. The inset panel zooms in during the first week and shows dotted lines with velocities¹⁷ of 613 km s^{-1} and $2,500 \text{ km s}^{-1}$ at a distance of 4.54 kpc. The apparent deceleration during the first week, along with the jump in size during the last week, can be explained by a two-component model consisting of a circular core surrounded by a ring where the flux ratio changes over time. The grey region shows the expansion rate of the core (lower edge) and the ring (upper edge). The solid line shows the effective size obtained by fitting the two-component model visibilities as a single circular disk. Error bars represent 1σ uncertainties computed from a bootstrap analysis.

¹The CHARA Array of Georgia State University, Mount Wilson Observatory, Mount Wilson, California 91023, USA. ²Center for High Angular Resolution Astronomy and Department of Physics and Astronomy, Georgia State University, PO Box 5060, Atlanta, Georgia 30302, USA. ³Laboratoire Lagrange, UMR 7293, Université de Sophia-Antipolis (UNS)–Centre National de la Recherche Scientifique (CNRS)–Observatoire de la Côte d’Azur (OCA), Boulevard de l’Observatoire, CS 34229, F-06304 Nice, Cedex 4, France. ⁴Department of Astronomy, University of Michigan, 941 Dennison Building, Ann Arbor, Michigan 48109, USA. ⁵National Optical Astronomy Observatory, 950 North Cherry Avenue, Tucson, Arizona 85719, USA. ⁶Université de Lyon; Université Lyon 1, Observatoire de Lyon, 9 avenue Charles André, 69230 Saint Genis Laval; CNRS UMR 5574, Centre de Recherche Astrophysique de Lyon; École Normale Supérieure, 69007 Lyon, France. ⁷Department of Astronomy, Yale University, New Haven, Connecticut 06511, USA. ⁸Sydney Institute for Astronomy, School of Physics, University of Sydney, New South Wales 2006, Sydney, Australia. ⁹Research School of Astronomy and Astrophysics, Australian National University, Canberra, Australian Capital Territory 2611, Australia. ¹⁰Remote Sensing Division, Naval Research Laboratory, 4555 Overlook Avenue Southwest, Washington, DC 20375, USA. ¹¹Département de Physique and Centre de Recherche en Astrophysique du Québec (CRAQ), Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada. ¹²Lowell Observatory, 1400 West Mars Hill Road, Flagstaff, Arizona 86001, USA. ¹³Max Planck Institute for Astronomy (MPIA), Königstuhl 17, 69117 Heidelberg, Germany. ¹⁴United States Naval Observatory, Flagstaff Station, 10391 West Naval Observatory Road, Flagstaff, Arizona 86001, USA. ¹⁵Astronomy and Astrophysics Division, Physical Research Laboratory, Navrangpura, Ahmedabad, Gujarat 380009, India. ¹⁶Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, California 91106, USA. ¹⁷Department of Astronomy, Boston University, Boston, Massachusetts 02215, USA.

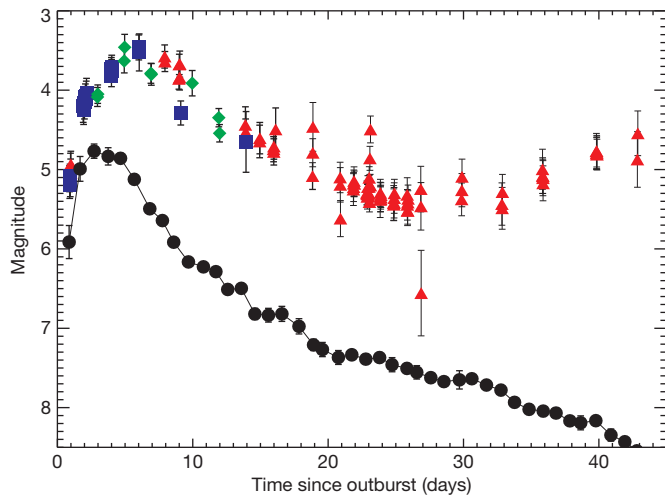


Figure 2 | Infrared light curve of Nova Del 2013. The magnitudes and error bars were computed from the mean and standard deviation of counts recorded on the detector during each interferometric observation collected at the CHARA Array. The data were obtained in the H-band using two different beam combiners (blue squares and green diamonds) and in the K-band (red triangles). For comparison, we plot daily averages of photometric measurements in the V-band downloaded from the American Association of Variable Star Observers (AAVSO; black circles). The rise to peak brightness was slower in the infrared compared to the visible.

brightnesses. Therefore, at the earliest times, the nova can be approximated by a single uniform disk component where the optically thick pseudo-photosphere extends to the outermost, fastest-moving layers of the ejecta. After the peak in the visible light curve, the amount of flux in the ring drops, which indicates that the outer layers became optically thin and the pseudo-photosphere moved towards the inner, slower-moving, and denser layers. Therefore, the apparent deceleration in the expansion curve during the first few nights is probably caused by a diminishing contribution from the outer layers. Over the next few weeks, the absolute flux of both components decreased; however, the flux from the core dropped at a steeper rate compared with the ring. Therefore, the percentage of light in the ring increased relative to the core. These changes are consistent with the spectral evolution¹⁷ of Nova Del 2013

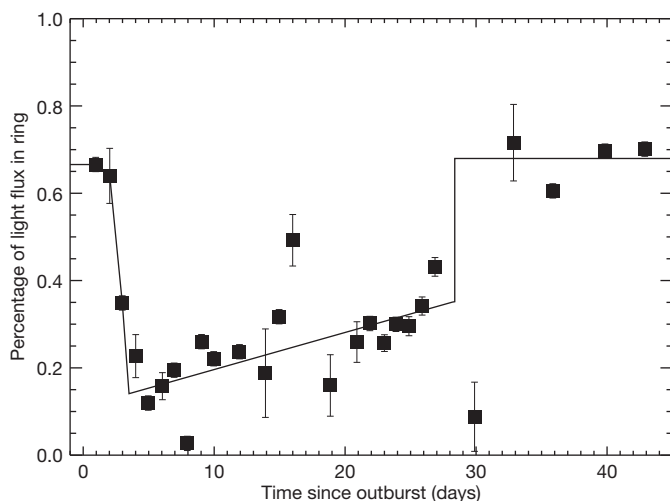


Figure 3 | Changes in the flux ratio of the two-component model. The solid line shows a fit where the ratio changes smoothly over time. For the first three nights, the ring represents the outer boundary of the optically thick pseudo-photosphere. After the peak in the visible light curve, the flux in the ring drops. On nights 4–27, the percentage of light from the ring increases at an approximately linear rate. On the last four nights, the ring contributes an average of 68% of the total light. Error bars represent 1σ uncertainties derived from a least-squares fit to the visibility data.

(see <http://www.astrosurf.com/aras/novae/Nova2013Del.html>). Optical spectra initially showed P Cygni absorption features during the optically thick fireball stage, which disappeared after about five days. Afterwards, strong emission lines began to dominate the spectrum, which are thought to form in the outer, optically thin layers². The rising strength of the emission lines relative to the continuum was also seen in spectra at infrared wavelengths (Extended Data Fig. 1). In the two-component model, we assumed a constant ratio between the size of the core and the halo. In reality, changes in the optical depth of the ejecta are likely to be more complex and could be investigated further by fitting a physical model^{18,19} simultaneously to interferometric and spectroscopic data. Understanding how the optical depth changes within the expanding ejecta affects the physical interpretation of novae light curves²⁰.

During the last five nights, the two-component model shows an enhancement in the flux of the outer layers, with the ring contributing about 68% of the total light. This coincides with an increase in the infrared flux from the nova²¹ (see Fig. 2). This suggests the formation of dust in the outer layers, perhaps in denser clumps within the ejecta², that contributes thermal emission. Alternatively, as the ejecta expands, a recombination front might propagate through the material²². Near-infrared free-free emission from ionized gas in the inner regions would decline and shift the effective size boundary to the outer layers. This interpretation is consistent with the larger size of the effective photosphere at later dates, but it does not explain the brightening in the near-infrared light curve. The effective size of our two-component model is shown as the solid line in Fig. 1; the model reproduces the changes in slope of the expansion curve.

We reconstructed model-independent images of Nova Del 2013 using the Markov chain imager²³ (MACIM) for data obtained on three nights during the first week with sufficient baseline coverage. As shown in Fig. 4, the images show a striking similarity to the two-component model, with an optically thick core surrounded by a halo of fainter emission.

Combining the angular expansion rate with the radial velocity of the ejected material provides a way of measuring the distance² to Nova Del 2013. The velocities reported in the literature¹⁷ range from -600 km s^{-1} to $-2,500 \text{ km s}^{-1}$, consistent with the expectation that lines form at different layers in the expanding atmosphere with a radial outflow velocity proportional to the distance from the white dwarf explosion site. We selected the Si II wavelength $\lambda = 6,347 \text{ \AA}$ and $\lambda = 6,371 \text{ \AA}$ absorption lines observed during the first week to represent the deeper layers, because they probably form in the dense, cooler gas immediately above the continuum-forming layer. From an analysis of spectra downloaded

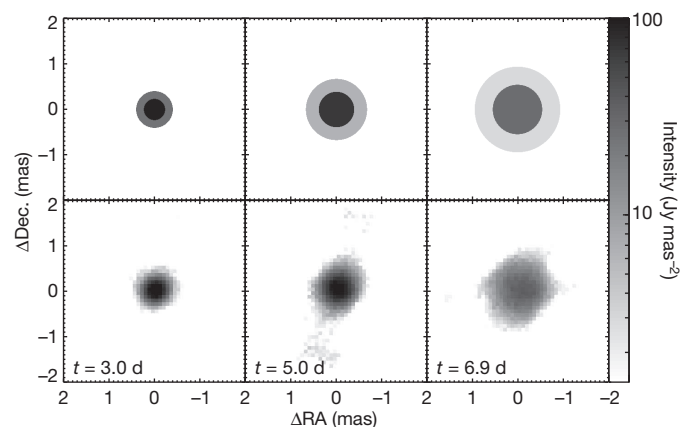


Figure 4 | Model and reconstructed images of Nova Del 2013. The two-component model consists of a circular core surrounded by a fainter ring (top row). The images reconstructed using MACIM show a similar spatial structure (bottom row). The data used in the reconstructions were obtained with the Michigan Infrared Combiner (MIRC) at the CHARA Array on UT 2013 August 17, 19 and 21 ($t = 3.0$ days, 5.0 days and 6.9 days after the outburst) and have good baseline coverage on the sky. We scaled the flux by the infrared H-band magnitude measured on each night to show how the surface brightness changes.

from the archive of the Astronomical Ring for Access to Spectroscopy¹⁷, we estimated the outflow speed near the continuum-forming layer to be $V_{\text{ejection}} = 613 \pm 79 \text{ km s}^{-1}$. Combining this velocity with the angular expansion rate, we derived a distance of $4.54 \pm 0.59 \text{ kpc}$ to Nova Del 2013. Empirical evidence suggests that novae have prolate spheroid²⁴ or bipolar¹⁹ shapes, so the true distance will depend on its geometry and inclination in the plane of the sky²⁵. Assuming a probable inclination of 45° for Nova Del 2013 (S. Shore, personal communication, 16 April 2014), we estimate that these effects contribute to a systematic difference in the distance of less than 5%. Using the light curve decay times and the reddening reported¹⁵ for Nova Del 2013 (assuming an $R_V = 3.1$ extinction law), the maximum magnitude to rate of decline (MMRD) relation²⁶ gives a distance ranging from 3.3 kpc to 4.1 kpc, depending on the adoption of the linear or nonlinear MMRD formulations. The expansion parallax also confirms the distance of 4.2 kpc based on a spectral comparison²⁷ to the nova OS Andromedae 1986.

To investigate the development of asymmetric features in Nova Del 2013, we fitted a uniformly bright ellipse to the visibilities on nights with coverage over a range of position angles on the sky (from two to nine days after the outburst; Extended Data Table 2). For each night, the ellipse produces a significantly lower χ^2 compared with the circular disk ($P < 0.01$ for most nights; see Methods). The nova appears to be around 13% larger along the major axis compared with the minor axis, suggesting that the ejecta could be bipolar. For the remaining nights with limited sky coverage, we fixed the axis ratio and position angle to their average values ($\theta_{\text{major}}/\theta_{\text{minor}} = 1.13 \pm 0.07$, $\text{PA} = 128^\circ \pm 30^\circ$) and solved for the size of the ellipse. On average, we find that the circular disk diameters underpredict the mean angular diameter of the ellipse by about 2%. Therefore, the assumption of circular symmetry in the two-component model should not strongly bias our results.

In addition to the visibility amplitudes that provide information on the size and shape of the source, interferometric closure phases obtained when combining the light from three or more telescopes indicate deviations from point symmetry (whether the brightness is the same when reflected through a point at the centre of the source distribution). The closure phases from the first two sets of data obtained with all six telescopes ($t = 3.0$ and 5.0 days) are consistent with zero within the uncertainties, suggesting that the source was point-symmetric within the resolution limits. By the third set of observations with all six telescopes ($t = 6.9$ days; see Fig. 4), the closure phases rise steadily to about 60° (Extended Data Fig. 2), suggesting the detection of a point asymmetry in the brightness distribution. If the outflow is bipolar, as suggested by our elliptical fits, then this could indicate a difference in the brightness between the two lobes (perhaps due to viewing angle) or, alternatively, the development of clumpy structures within the expanding material.

The ellipticity and closure phase asymmetries indicate that non-spherical structures developed as early as a few days after the outburst. Along with results from previous interferometric studies of novae^{28–30}, this suggests that novae explosions might be inherently bipolar or that the elliptical shape develops early during the common envelope phase.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 May; accepted 3 September 2014.

Published online 26 October 2014.

1. Bode, M. F. & Evans, A. *Classical Novae* (Cambridge Univ. Press, 2008).
2. Gehrz, R. D. The infrared temporal development of classical novae. *Annu. Rev. Astron. Astrophys.* **26**, 377–412 (1988).
3. Slavin, A. J., O'Brien, T. J. & Dunlop, J. S. A deep optical imaging study of the nebular remnants of classical novae. *Mon. Not. R. Astron. Soc.* **276**, 353–371 (1995).
4. Woudt, P. A. *et al.* The expanding bipolar shell of the helium nova V445 Puppi. *Astrophys. J.* **706**, 738–746 (2009).
5. Chesneau, O. *et al.* The expanding dusty bipolar nebula around the nova V1280 Scorpi. *Astron. Astrophys.* **545**, A63 (2012).
6. Livio, M., Shankar, A., Burkert, A. & Truran, J. W. The common envelope phase in the outbursts of classical novae. *Astrophys. J.* **356**, 250–254 (1990).

7. Lloyd, H. M., O'Brien, T. J. & Bode, M. F. Shaping of nova remnants by binary motion. *Mon. Not. R. Astron. Soc.* **284**, 137–147 (1997).
8. Orio, M. & Shaviv, G. Local thermonuclear runaways among classical novae. *Astrophys. Space Sci.* **202**, 273–288 (1993).
9. Porter, J. M., O'Brien, T. J. & Bode, M. F. On the asphericity of nova remnants caused by rotating white dwarf envelopes. *Mon. Not. R. Astron. Soc.* **296**, 943–948 (1998).
10. Scott, A. D. Nova outbursts on rotating oblate white dwarfs. *Mon. Not. R. Astron. Soc.* **313**, 775–782 (2000).
11. Chesneau, O. & Banerjee, D. P. K. Interferometric studies of novae in the infrared. *Bull. Astron. Soc. India* **40**, 267–292 (2012).
12. Nakano, S. *et al.* Nova Delphini 2013 = Pnv J20233073+2046041. *Central Bureau Electron. Teleg.* **3628**, 1 (2013).
13. ten Brummelaar, T. A. *et al.* First results from the CHARA array. II. A description of the instrument. *Astrophys. J.* **628**, 453–465 (2005).
14. Wren, J., Vestrand, W. T., Wozniak, P. & Davis, H. Pre-discovery detections of the Nova Delphini outburst. *Astron. Teleg.* **5316**, 1 (2013).
15. Munari, U., Henden, A., Dallaporta, S. & Cherini, G. Photometric evolution of Nova Del 2013 (V339 Del) during the optically thick phase. *Inform. Bull. Variable Stars* **6080**, 1 (2013).
16. Deacon, N. R. *et al.* Pre-outburst observations of Nova Del 2013 from Pan-STARRS 1. *Astron. Astrophys.* **563**, A129 (2014).
17. Shore, S. N. *et al.* Continuing spectroscopic observations (3500–8800 Å) of Nova Del 2013 with the Ondrejov Observatory and the ARAS group. *Astron. Teleg.* **5312**, 1 (2013).
18. Gill, C. D. & O'Brien, T. J. Emission-line profiles from model nova shells. *Mon. Not. R. Astron. Soc.* **307**, 677–684 (1999).
19. Shore, S. N. *et al.* The spectroscopic evolution of the recurrent nova T Pyxidis during its 2011 outburst. II. The optically thin phase and the structure of the ejecta in recurrent novae. *Astron. Astrophys.* **549**, A140 (2013).
20. Shore, S. N. A unified model for the spectrophotometric development of classical and recurrent novae. The role of asphericity of the ejecta. *Astron. Astrophys.* **559**, L7 (2013).
21. Taranova, O. G., Tatarnikov, A. M., Shenavrin, V. I. & Tatarnikova, A. A. Infrared photometry of Nova Delphini 2013 (=V339 Del) in the first sixty days after its outburst. *Astron. Lett.* **40**, 120–124 (2014).
22. Shore, S. N., Augusteijn, T., Ederocite, A. & Uthas, H. The spectroscopic evolution of the recurrent nova T Pyxidis during its 2011 outburst. I. The optically thick phase and the origin of moving lines in novae. *Astron. Astrophys.* **533**, L8 (2011).
23. Ireland, M. J., Monnier, J. D. & Thureau, N. Monte-Carlo imaging for optical interferometry. *Proc. SPIE* **6268**, 62681T (2006).
24. Ford, H. C. & Ciardullo, R. in *The Extragalactic Distance Scale* (eds van den Bergh, S. & Pritchet, C. J.) *ASP Conf. Ser.* **4**, 128–144 (1988).
25. Wade, R. A., Harlow, J. J. B. & Ciardullo, R. Biases in expansion distances of novae arising from the prolate geometry of nova shells. *Publ. Astron. Soc. Pacif.* **112**, 614–624 (2000).
26. Downes, R. A. & Duerbeck, H. W. Optical imaging of nova shells and the maximum magnitude-rate of decline relationship. *Astron. J.* **120**, 2007–2037 (2000).
27. Shore, S. N. *et al.* First high resolution ultraviolet (HST/STIS) and supporting optical spectroscopy of V339 Del = Nova Del 2013. *Astron. Teleg.* **5409**, 1 (2013); correction **5410**, 1 (2013).
28. Lane, B. F. *et al.* Interferometric observations of V1663 Aquilae (Nova Aql 2005). *Astrophys. J.* **669**, 1150–1155 (2007).
29. Chesneau, O. *et al.* AMBER/VLTI interferometric observations of the recurrent Nova RS Ophiuchi 5.5 days after outburst. *Astron. Astrophys.* **464**, 119–126 (2007).
30. Chesneau, O. *et al.* The 2011 outburst of the recurrent nova T Pyxidis. Evidence for a face-on bipolar ejection. *Astron. Astrophys.* **534**, L11 (2011).

Acknowledgements We acknowledge the variable star observations from the AAVSO International Database contributed by observers worldwide and used in this research. We thank O. Garde and other members of the Astronomical Ring for Access to Spectroscopy for use of their archive of Nova Del 2013 spectra. We thank G. J. Schwarz, S. N. Shore, and F. M. Walter for discussions that helped us to interpret the nova observations. This material is based upon work supported by the National Science Foundation under grant number AST-1009080. The CHARA Array is funded by the National Science Foundation through NSF grants AST 0908253 and AST 1211129, and by Georgia State University through the College of Arts and Sciences. This publication made use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

Author Contributions Observations with the CHARA Array were originally proposed by B.K. and D.R.G. Modelling and interpreting the angular expansion curve and asymmetries were done by G.H.S., D.R.G., B.K., T.t.B., O.C., I.T.-B. and S.T.R. The CHARA data were reduced by T.t.B., J.D.M., O.C., I.T.-B., D.M., V.M., C.D.F., N.S. and G.H.S. The observations were planned and conducted by C.D.F., N.S., N.V., B.K., D.R.G., T.B., G.H.S., D.M., A.M., N.N., P.S., M.I., V.M., P.T., J.J., N.D.R., R.M.R., G.v.B., K.v.B. and R.T.Z. Observational setup and technical support were provided by J.S., L.S., N.H.T. and X.C. Administrative oversight and access to CHARA were provided by H.A.M. and T.t.B. Reconstructing and interpreting the nova images were done by G.H.S., F.B., J.D.M. and B.K. Infrared magnitudes derived from CHARA data were computed by N.S. and G.H.S. Infrared spectra were taken and reduced by D.P.K.B., N.M.A., V.J., P.S.M., J.B. and analysed by D.R.G. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.H.S. (schaefer@chara-array.org).

METHODS

Interferometric observations with the CHARA Array. The CHARA Array¹³ is an optical/infrared interferometer located on Mount Wilson, California. The array has six 1-m telescopes arranged in a Y configuration with baselines ranging from 34 m to 331 m. The longest baseline provides a spatial resolution of 0.5 mas at a wavelength of 1.6 μm . We observed Nova Del 2013 in the infrared part of the spectrum using the CLASSIC two-beam combiner³¹, the CLIMB three-beam combiner³¹, and the MIRC six-beam combiner³². The data were reduced using the standard reduction pipeline for each instrument. Data products include the squared, normalized visibility amplitude of the interference fringes for the two-baseline combiners and the squared, normalized visibility amplitudes and closure phases for the instruments that combine the light from three or more telescopes (see Extended Data Figs 2–4). Before and after each observation, we observed unresolved stars or stars with known angular diameters^{33–35} to calibrate the interferometric measurements. Extended Data Table 1 lists the UT date, modified Julian Day (MJD), time since the outburst ($t_0 = \text{MJD } 56518.277$), beam combiner, telescope configuration, filter and calibrators used during each observation. The calibrated data will be available through the Optical Interferometry Database developed by the Jean-Marie Mariotti Center (<http://www.jmmc.fr/oidb.htm>). On five nights we also collected data using two visible light beam combiners; these data will be discussed in a subsequent publication (O.C., D.M., P.T., V.M., I.T.-B., D.P.K.B., E. Lagadec, M.I., N.N., A.M., P.S., R.T.Z., G.H.S., T.t.B., D.R.G., H.A.M., T.B., C.D.F., N.V., N.S., J.S. and L.S., manuscript in preparation).

In addition to the interferometric measurements, we also computed the infrared magnitudes of Nova Del 2013 using the number of counts recorded on the detector. We compared the counts from Nova Del 2013 with the calibrator stars observed immediately before and after. We calibrated the photometry based on the 2MASS magnitudes³⁶ of the calibrators. We present the photometry in Extended Data Table 3. Figure 2 shows the infrared light curve of Nova Del 2013. For comparison, we also plot daily averages of V-band photometric measurements downloaded from AAVSO³⁷. The rise to peak brightness was slower in the infrared compared to the visible part of the spectrum. Based on a second-order polynomial fit to the H-band and K'-band measurements near the maximum (<4.3 mag), the near-infrared light curve reached a maximum of 3.61 mag at MJD 56524.6 ($t = 6.3$ days). The maximum brightness in the visible¹⁵ occurred at an earlier time of MJD 56520.4 ($t = 2.1$ days) at $V \approx 4.46$ mag.

Correcting for the effects of line emission. The emergence of strong emission lines in the infrared spectrum of Nova Del 2013 (such as Br γ in the K-band) will broaden the central envelope of the interference fringe and create sidelobes farther out³⁸. We investigated the effects of the emission lines on the calibration of the interferometric data obtained with the CLASSIC and CLIMB beam combiners. The reduction code outputs the mean power spectrum for each target. We fitted a rectangular top-hat function to each power spectrum and calculated the mean centre and width for the observations on each night. By comparing the results from the nova with the calibrators, we found that the central wavelength of the filter remained constant over time. We also found that the effective bandwidth of the filter remained the same during the first two weeks and then decreased by about $14\% \pm 3\%$ as the emission lines appeared in the nova spectra (see Extended Data Table 4). Essentially, the emission lines put a larger fraction of the flux at the centre of the wavelength region, thereby effectively decreasing the full-width at half-maximum (FWHM) of the filter. The CLASSIC/CLIMB reduction software assumes a constant top-hat function for the width of the filter. Therefore, when the effective bandwidth decreases, the visibility is overestimated and the angular size of the nova is underestimated.

To correct for this effect, we multiplied the CLASSIC/CLIMB visibilities by the ratio of the mean effective width of the nova observations relative to the calibrators measured on each night. We added the uncertainty in the bandwidth ratio in quadrature with the uncertainties in the visibilities reported by the reduction code. We applied this correction starting on UT 2013 August 28, when the effects of the emission lines became measurable. During the first two weeks, when the contribution from the emission lines had a negligible effect on the calibration, we assumed the default bandwidth used in the reduction code, but included the average uncertainty in the nova-to-calibrator bandwidth ratio of ± 0.164 in the error budget so that all of the CLASSIC and CLIMB visibilities would have similar weights during the global two-component fit. For the uniform disk fit to the data from each night, the bandwidth correction led to the measurement of the angular size of the nova being larger at the approximately 5% level after the emission lines developed.

Uniform disk and uniform ellipse fits. For each individual night of CHARA data, we measured the angular diameter of Nova Del 2013 by fitting a uniformly bright, circular disk to the visibilities and performing a Levenberg–Marquardt least-squares minimization using the IDL *mpfit* package³⁹ (<http://purl.com/net/mpfit>). The uniform disk diameters (θ_{UD}), reduced χ^2_r (where ν is the number of degrees of freedom), and number of visibility measurements $N(V^2)$ on each night are listed in the last three columns of Extended Data Table 1.

For five nights with sufficient baseline coverage over a range of position angles, we fitted a uniformly bright ellipse to the visibility data. The fitted parameters include the size of the major and minor axes and position angle of the major axis (θ_{major} , θ_{minor} , PA); the values are given in Extended Data Table 2. For four of the five nights, performing an F -test indicates that the uniform ellipse produces a significant improvement in χ^2 compared with a uniform circle, at a significance level of $P = 0.01$. On UT 2013 August 16, the improvement is only at the 0.10 significance level.

For both the uniform disk and uniform ellipse fits, we determined uncertainties in the parameters using a bootstrap approach⁴⁰. In each iteration, we created a synthetic sample of visibility measurements from our original set of measurements by randomly selecting the same total number of visibilities, with replacement. Therefore, in the synthetic sample, some measurements are repeated and others are left out. We applied normally distributed uncertainties to the synthetic measurements (using the measured uncertainties from the original data). We then fitted a uniform circle or uniform ellipse to the synthetic data set to determine the best-fitting parameters for that iteration. We performed 1,000 iterations and computed uncertainties based on the standard deviation of the parameter distributions. These uncertainties are reported in Extended Data Tables 1 and 2.

We expected the size of the nova to vary over the course of the night. The maximum length of observing time was about 6 h on UT 2013 August 16 and September 6. Given an expansion rate of 0.156 mas per day, we would expect the size of the nova to change by around 0.04 mas over the course of the observations. For the three nights with the smallest uncertainties on the angular diameter measurements (± 0.005 mas on UT 2013 August 17, 19 and 21), the length of observing time ranged from 30 min to 45 min, so the change in size is expected to be <0.005 mas. For all other nights, the length of observing time varied from about 30 min to 3 h, corresponding to a change in size of 0.003 mas to 0.02 mas. For these nights, the quoted uncertainties on the angular diameters in Extended Data Table 1 are larger than ± 0.02 mas, which should account for any changes in size over the course of the night. For the two-component model described in the next section, we fitted a dynamic model directly to the visibilities versus time, so it accounts for changes in size over the course of the night.

Two-component model. The two-component model consists of a uniformly bright, circular disk surrounded by a uniformly bright ring. The angular diameter of the central uniform disk θ_{UD} and the outer diameter of the ring θ_{ring} are given by:

$$\theta_{\text{UD}}(t) = \frac{d\theta}{dt}(t - t_0) \quad (1)$$

$$\theta_{\text{ring}}(t) = C_{\text{ring}}\theta_{\text{UD}}(t) \quad (2)$$

where $\frac{d\theta}{dt}$ is the angular expansion rate of the central core and C_{ring} is the ratio of the outer ring diameter relative to the uniform disk diameter.

We determined the two global parameters ($\frac{d\theta}{dt}$ and C_{ring}) by minimizing the cumulative χ^2 between the measured and model visibilities across all nights. To do this, we searched through a grid of values for the expansion rate and size ratio and calculated the size of each component during the times of observation. Then, for each night of observation, we performed a Levenberg–Marquardt least-squares minimization using the IDL *mpfit* package³⁹ to solve for the best-fitting flux ratio between the ring and the core component.

There are degeneracies between the overall size of the nova and the flux ratio of the two components; a smaller expansion rate can be compensated by placing more flux into the outer component. Therefore, we followed a two-step procedure in fitting the expanding two-component model simultaneously to all nights. Initially, we allowed the flux in the ring relative to the core to vary each night, but with a limit set so that the surface brightness of the ring would not exceed that of the core. The initial restriction prevented a large increase in the fitted brightness of the ring at later nights. Using the grid-search described above, we determined an expansion rate of 0.156 ± 0.003 mas per day for the core diameter and a size ratio of 1.73 ± 0.02 between the outer ring diameter and the core diameter. We then fixed the expansion rate and size ratio and solved for the ring-to-core flux ratio for each night after removing the restriction on the surface brightness on all but the first three nights (we would not expect a clearing of the central region at such early times). This allows for a moderate brightening of the ring at later times.

In Fig. 1, the solid line shows the effective angular size of the two-component model. This curve was generated by assuming that the percentage of light from the ring varies smoothly over time (solid line in Fig. 3) and computing the model visibilities for each night based on the expansion rate and size ratio. We then fitted a single, uniform, circular disk to the model visibilities from each night to compute the effective size. Extended Data Fig. 5 shows how the size and flux ratios of the two components change during the nights of observation.

Emission line flux from infrared spectroscopy. To investigate how the emission line flux changed in the infrared, we used low dispersion spectra obtained with the TripleSPEC spectrograph^{41,42} on the 200-inch Hale Telescope at Palomar Observatory on UT 2013 August 20 and 23. We reduced the spectra using the Spextool reduction program modified for TripleSpec⁴³, which includes routines to correct for detector artefacts⁴⁴. We calibrated and removed the telluric lines in the nova spectrum using the *xtellcor* program⁴⁵, which compares a high signal-to-noise spectrum of Vega to observations of a nearby A0 star, SAO 88500, taken immediately after each observation of the nova. We also used infrared spectra⁴⁶ obtained at Mount Abu, India⁴⁷, on UT 2013 August 28 and 29, and September 8 and 20.

We collected these spectra into two regions covering the wavelength range of the CHARA CLASSIC/CLIMB H-band and K'-band filters and normalized the flux to unity at 1.65 μm and 2.2 μm , respectively. The H-band is dominated by the upper-level transitions of the hydrogen Brackett series plus a number of C I lines, and the K'-band is dominated by Br γ , He I $\lambda = 2.0585 \mu\text{m}$ plus some C I lines⁴⁷.

To estimate the flux from the continuum, we made spline fits to the lower envelope of each spectrum. We then summed both the continuum flux and the emission flux (total minus continuum) across the wavelength bands to form a ratio of the integrated emission line flux relative to the integrated continuum flux. As shown in Extended Data Fig. 1, the emission strength grew over the time frame of the CHARA observations of Nova Del 2013 and began to decrease in the final spectrum. These changes are similar to the changes we measured in the effective bandwidth of the interferometric observations.

The strengthening in the emission lines helps to explain why the expansion curve of Nova Del 2013 became steeper after the first week. The emission lines form over a greater radial extent in the expanding envelope than does the continuum from the pseudo-photosphere. Thus, if the outflow follows a Hubble-type relation with the velocity proportional to radial distance⁴⁸, then the emission lines sample parts of the envelope that were expanding faster than the photosphere. We interpolated linearly between the spectroscopic measurements to obtain the ratio of the emission line flux to the continuum flux at the times of the CHARA observations and used this as the initial estimate of the flux ratio in the halo ring relative to the uniform disk core in the two-component fit.

The drop in the emission-to-continuum flux ratio after the fourth week can be attributed to the onset of dust formation. This can be seen in the K-band spectra, where the continuum slopes down in the first spectra and becomes flat in the last spectrum. This is consistent with flux from a dusty component that contributes more to the continuum and increases with wavelength. Our two-component model based on the interferometric observations suggests that if dust formation occurred, that it happened over a larger spatial scale than that of the expanding pseudo-photosphere.

We also made measurements of the flux-weighted central wavelength of each infrared spectrum. These were essentially constant to within a per cent or so during these observations, consistent with our results from the analysis of the fringe power spectra.

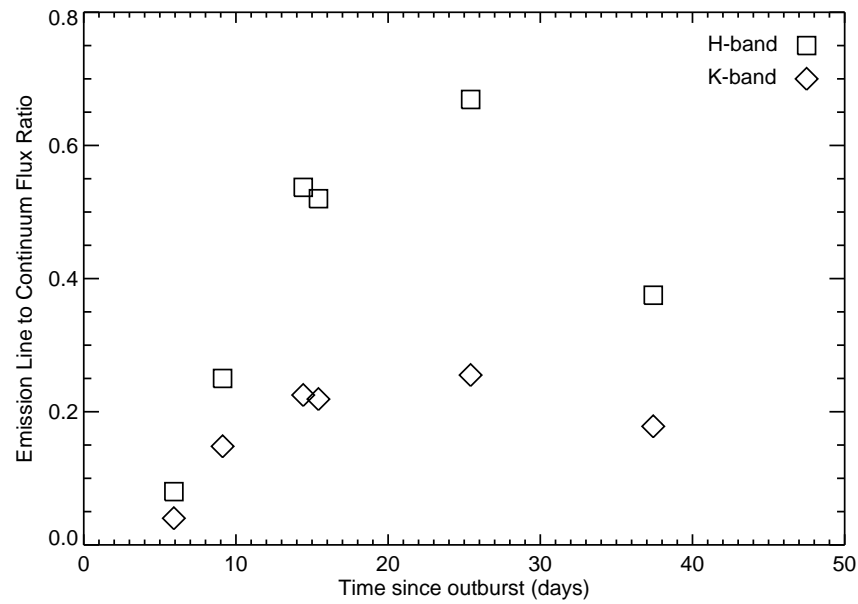
Measuring the radial velocity of the optically thick core. To compare the angular expansion rate with the radial velocity of the nova outflow, we needed to decide which features in the optical spectra are most representative of the kinematics of the near-infrared continuum. The velocities reported in the literature¹⁷ range from -600 km s^{-1} to $-2,500 \text{ km s}^{-1}$, consistent with the expectation that lines form at different layers in the expanding atmosphere with a radial outflow velocity proportional to distance from the white dwarf explosion site. The optical depth unity boundary for the near-infrared continuum must form at higher densities deep inside the outflow (compared to the P Cygni lines and emission lines that form further out in lower density gas). The absorption lines observed during the first week are probably the most representative of the deeper layers, because they form in the dense, cooler gas immediately above the continuum forming layer. We selected the Si II $\lambda = 6,347 \text{ \AA}$ and $6,371 \text{ \AA}$ absorption lines for measurement because they were observed throughout the first week and are relatively free of blending and interference from Earth's atmospheric lines and nova emission lines.

We downloaded six spectra with a high resolving power ($R = 10,000$) obtained over the first week by O. Garde from the archive of the Astronomical Ring for Access

to Spectroscopy¹⁷ and transformed them to a unit continuum on a heliocentric wavelength grid. We measured radial velocities for the Si II lines by cross-correlation with a model spectrum ($T_{\text{eff}} = 8,000 \text{ K}$, $\log g = 2.0$, solar metallicity) based upon ATLAS atmospheres from R. Kurucz. The mean velocity and standard deviation from this sample (MJD 56519.8 – 56524.8) are $-598 \pm 51 \text{ km s}^{-1}$. The local standard of rest radial velocity at the probable distance of the nova is $+15 \text{ km s}^{-1}$, so the implied outflow velocity is $-613 \pm 51 \text{ km s}^{-1}$. However, we do not know the peculiar velocity of the nova relative to its local standard of rest, so we applied a representative velocity dispersion of 60 km s^{-1} for white dwarfs to revise the uncertainty in our final estimate of the outflow speed near the continuum forming layer, $V_{\text{ejection}} = 613 \pm 79 \text{ km s}^{-1}$.

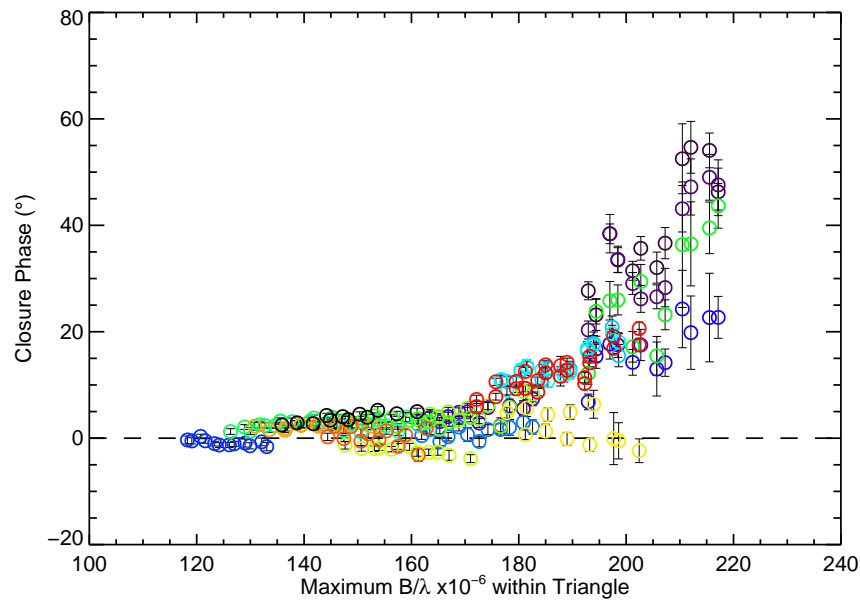
We note that this is strictly an upper limit for the velocity of the continuum forming layers because the absorption lines form above the continuum in a somewhat faster outflow. Additionally, the estimate reflects the geometry of the outflow, which is dependent on the axial inclination for a bipolar ejection. For example, if the absorption lines formed in a spherically expanding photosphere, then we might expect to observe line boundaries ranging from zero (flux from the limb) to $-V_{\text{outflow}}$ (flux from the centre), with a mean velocity between these limits. However, the observed Si II lines are rather narrow (FWHM is 280 km s^{-1}), contrary to this prediction. On the other hand, for a bipolar outflow the absorption velocities may be much more restricted to the radial ejection speed along the line of sight. Consequently, until future observations reveal details of the geometry of the nova remnant, we will simply assume that the measured radial velocity is the same as the radial outflow speed in our direction and that this is also the same as the transverse expansion speed, as observed in the plane of the sky.

31. ten Brummelaar, T. A. *et al.* The CLASSIC/CLIMB Beam Combiner at the CHARA Array. *J. Astron. Instrum.* **2**, 1340004 (2013).
32. Monnier, J. D. *et al.* Michigan Infrared Combiner (MIRC): commissioning results at the CHARA Array. *Proc. SPIE* **6268**, 62681P (2006).
33. Bonneau, D. *et al.* SearchCal: a virtual observatory tool for searching calibrators in optical long baseline interferometry. I. The bright object case. *Astron. Astrophys.* **456**, 789 (2006).
34. Boyajian, T. S. *et al.* Stellar diameters and temperatures. III. Main-sequence A, F, G, and K stars: additional high-precision measurements and empirical relations. *Astrophys. J.* **771**, 40 (2013).
35. Crepp, J. R. *et al.* The dynamical mass and three-dimensional orbit of HR7672B: a benchmark brown dwarf with high eccentricity. *Astrophys. J.* **751**, 97 (2012).
36. Skrutskie, M. F. *et al.* The Two Micron All Sky Survey (2MASS). *Astron. J.* **131**, 1163–1183 (2006).
37. Henden, A. A. The AAVSO International Database. <http://www.aavso.org> (2013).
38. Tuthill, P. *et al.* Double-Fourier spatio-spectral decoding. In *Advances in Stellar Interferometry*, *SPIE Conf. Ser.* **6268**, 62680X (2006). <http://dx.doi.org/10.1117/12.672277>.
39. Markwardt, C. B. in *Astronomical Data Analysis Software and Systems XVIII*, *Astron. Soc. Pacif. Conf. Ser.* **411**, 251–254 (2009).
40. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C* (Cambridge Univ. Press, 1992).
41. Wilson, J. C. *et al.* in *Ground-based Instrumentation for Astronomy*, *SPIE Conf. Ser.* **5492**, 1295–1305 (2004); <http://dx.doi.org/10.1117/12.550925>.
42. Herter, T. L. *et al.* in *Ground-based and Airborne Instrumentation for Astronomy II*, *SPIE Conf. Ser.* **7014**, 70140X (2008); <http://dx.doi.org/10.1117/12.789660>.
43. Cushing, M. C., Vacca, W. D. & Rayner, J. T. Spextool: a spectral extraction package for SpeX, a 0.8–5.5 micron cross-dispersed spectrograph. *Proc. Astron. Soc. Pacif.* **116**, 362–376 (2004).
44. Muirhead, P. S. *et al.* Precise stellar radial velocities of an M dwarf with a Michelson interferometer and a medium-resolution near-infrared spectrograph. *Publ. Astron. Soc. Pacif.* **123**, 709–724 (2011).
45. Vacca, W. D., Cushing, M. C. & Rayner, J. T. A method of correcting near-infrared spectra for telluric absorption. *Publ. Astron. Soc. Pacif.* **115**, 389–409 (2003).
46. Banerjee, D. P. K., Ashok, N. M., Joshi, V. & Evans, A. Ongoing near-infrared observations of V339 Del (Nova Del 2013). *Astron. Telegr.* **5404**, 1 (2013).
47. Das, R. K., Banerjee, D. P. K., Ashok, N. M. & Chesneau, O. Near-infrared studies of V1280 Sco (Nova Scorpii 2007). *Mon. Not. R. Astron. Soc.* **391**, 1874–1886 (2008).
48. Hjelming, R. M., Wade, C. M., Vandenberg, N. R. & Newell, R. T. Radio emission from nova shells. *Astron. J.* **84**, 1619–1631 (1979).



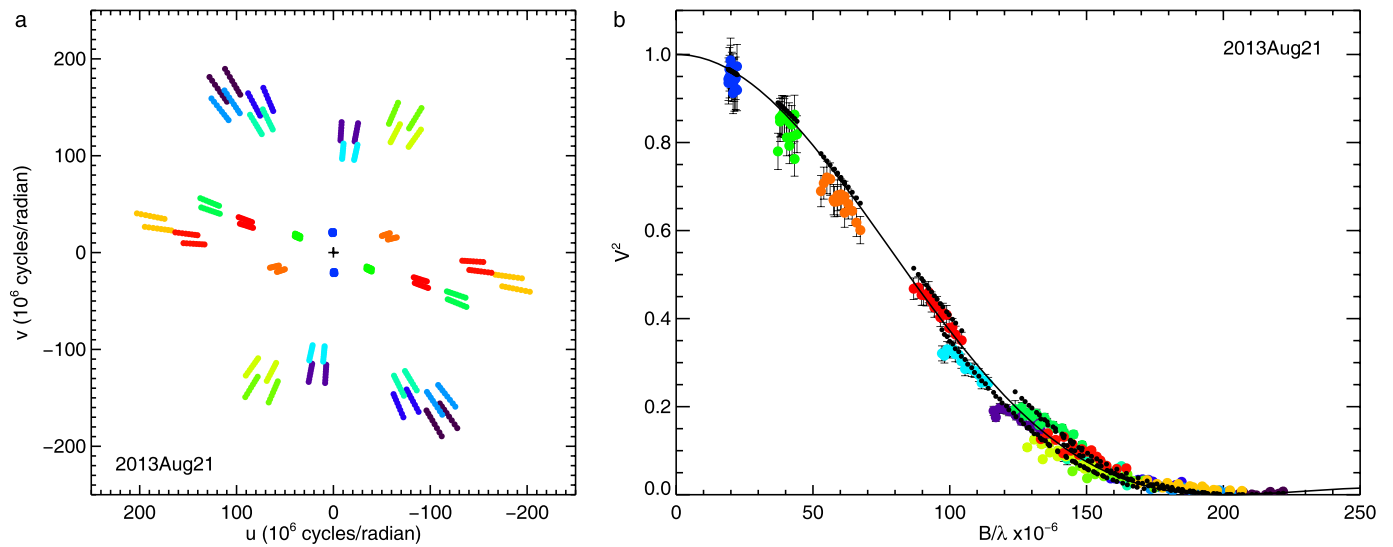
Extended Data Figure 1 | Integrated emission line to continuum flux ratios measured from infrared spectroscopy. The squares represent the H-band ratios while the diamonds represent the K'-band ratios. The rise in the emission line flux is consistent with an increasing contribution from optically thin

emission. The down turn in the curve during the last measurement is probably caused by a rising contribution of the continuum owing to the formation of dust.



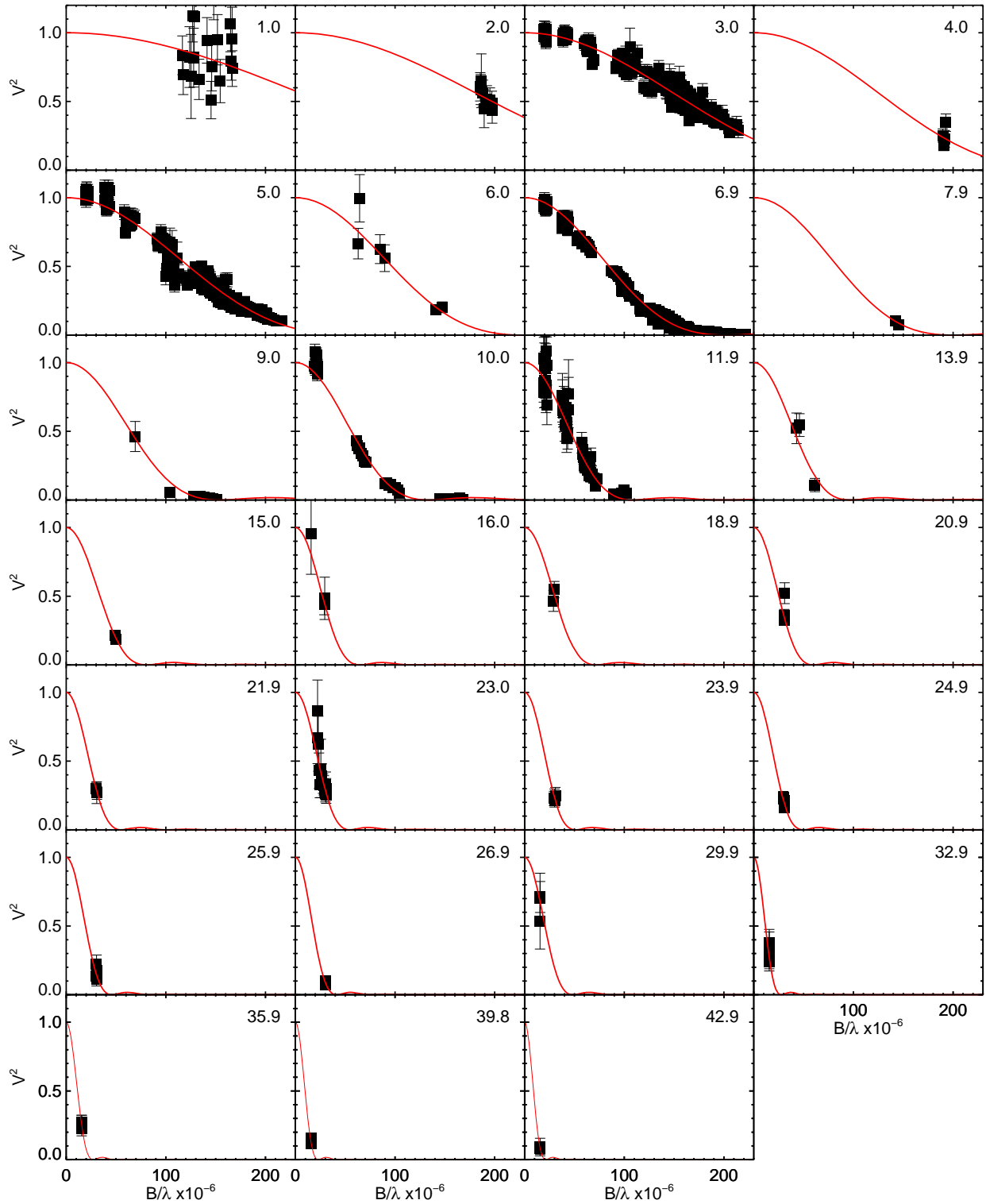
Extended Data Figure 2 | Closure phases measured with MIRC on UT 2013 August 21. The measurements are plotted against the length of the maximum baseline (B) for the group of three telescopes that were used to form the closure phase. The non-zero closure phases indicate an asymmetry in the light

distribution that is not point symmetric. The colours of the symbols are used to differentiate the measurements from each grouping of three telescopes in the array. The closure phases are measured in eight wavelength channels. Error bars represent 1σ measurement uncertainties.



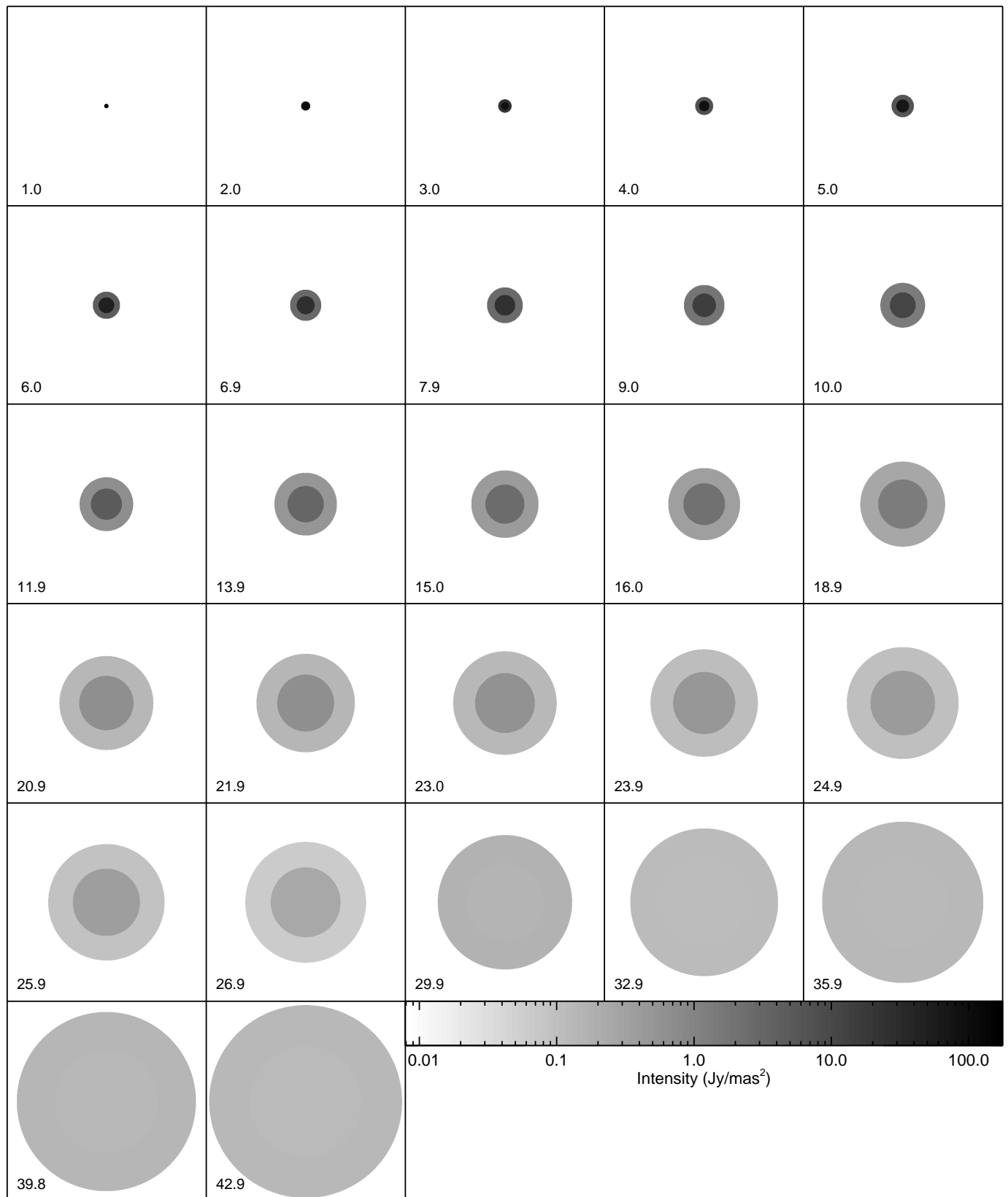
Extended Data Figure 3 | Example of MIRC observations obtained on UT 2013 August 21. The light is dispersed over eight wavelength channels in the H-band. **a**, Coverage of the interferometric baselines projected on the plane of the sky in right ascension (RA) and declination (Dec.) in units of spatial frequencies ($u = B_x/\lambda$, $v = B_y/\lambda$). **b**, Squared, normalized visibility amplitude

measurements, colour-coded to match the baselines on the left. The solid line shows the best-fitting uniform disk model. The small black dots show the best-fitting uniform ellipse model. Error bars represent 1σ measurement uncertainties.



Extended Data Figure 4 | Interferometric visibilities of Nova Del 2013 measured with the CHARA Array. The red line shows the best-fitting model for a uniformly bright, circular disk. The time since the explosion (in days) is

indicated in the upper right corner of each panel. The measurements were obtained with the CLASSIC, CLIMB, and MIRC beam combiners (see Extended Table 1). Error bars represent 1σ measurement uncertainties.



Extended Data Figure 5 | Time evolution of the two-component model of Nova Del 2013. The model consists of a circular core surrounded by a halo ring. The expansion rate and size ratio were determined by minimizing χ^2 across all of the nights while allowing the flux ratio to vary night by night. In plotting the images, we used the flux ratio measured directly for the first three nights, the linear fit plotted in Fig. 3 of the main paper for $t = 4$ –27 days, and

our measurement that the ring contributes an average of 68% of the light on the last five nights (dust emission in the outer layers). Each panel is 12 mas on a side. We scaled the model flux by the infrared magnitude measured on each night to show how the surface brightness changes. The time since the explosion (in days) is indicated in each panel. Intensity refers to the flux per unit area.

Extended Data Table 1 | Journal of observations and angular diameters measurements of Nova Del 2013

UT Date	MJD	t (days)	Comb*	Configuration	Filter [†]	Cal [‡]	θ_{UD} (mas)	χ^2_ν	N(V^2)
2013 Aug 15	56519.231	1.0	CL	S1E2W1	H, K'	1,2	0.414 ± 0.076	1.19	18
2013 Aug 16	56520.303	2.0	CL	S1E1	H	1,2	0.537 ± 0.028	0.23	9
2013 Aug 17	56521.260	3.0	M	W1S2S1E1E2W2	H	3	0.652 ± 0.005	1.40	180
2013 Aug 18	56522.284	4.0	C	S1E1	H	1	0.777 ± 0.038	1.77	4
2013 Aug 19	56523.248	5.0	M	W1S2S1E1E2W2	H	4,3	0.863 ± 0.004	4.10	180
2013 Aug 20	56524.325	6.0	CL	E2W2W1	H	1	1.096 ± 0.056	0.82	6
2013 Aug 21	56525.220	6.9	M	W1S2S1E1E2W2	H	4,3	1.250 ± 0.006	11.5	240
2013 Aug 22	56526.216	7.9	C	E1W1	K'	4	1.262 ± 0.046	0.17	2
2013 Aug 23	56527.310	9.0	CL	S1E1W1, S2E2W2	H, K'	1,4,5	1.643 ± 0.077	5.10	15
2013 Aug 24 [§]	56528.242	10.0	M	W1S2S1E1E2W2	H	3	1.885 ± 0.016	4.86	32
2013 Aug 26 [§]	56530.208	11.9	M	W1S2S1E1E2W2	H	3	2.314 ± 0.037	1.09	61
2013 Aug 28	56532.191	13.9	CL	W1W2	H, K'	4	2.62 ± 0.27	2.49	3
2013 Aug 29	56533.235	15.0	CL	W1W2	K'	4	3.133 ± 0.076	0.29	2
2013 Aug 30	56534.273	16.0	CL	E1E2, S1S2	K'	1,4	3.89 ± 0.72	0.19	3
2013 Sep 02	56537.146	18.9	C	E1E2	K'	4	3.51 ± 0.31	1.10	2
2013 Sep 04	56539.178	20.9	C	E1E2	K'	1,4	4.22 ± 0.22	3.32	3
2013 Sep 05	56540.164	21.9	C	E1E2	K'	1,4	4.57 ± 0.18	0.04	3
2013 Sep 06	56541.280	23.0	C	E1E2	K'	1,4	4.62 ± 0.13	0.40	16
2013 Sep 07	56542.158	23.9	C	E1E2	K'	1	4.94 ± 0.11	0.23	6
2013 Sep 08	56543.150	24.9	CL	E1E2	K'	1	5.11 ± 0.12	1.68	5
2013 Sep 09	56544.156	25.9	CL	E1E2	K'	1	5.49 ± 0.11	1.76	6
2013 Sep 10	56545.139	26.9	CL	E1E2	K'	1	6.09 ± 0.17	0.38	2
2013 Sep 13	56548.160	29.9	CL	S1S2	K'	1	5.19 ± 0.97	0.31	3
2013 Sep 16	56551.137	32.9	C	S1S2	K'	1	9.12 ± 0.74	0.85	3
2013 Sep 19	56554.138	35.9	C	S1S2	K'	1	9.33 ± 0.30	0.12	5
2013 Sep 23	56558.121	39.8	C	S1S2	K'	1	10.96 ± 0.18	0.52	5
2013 Sep 26	56561.159	42.9	C	S1S2	K'	1	11.83 ± 0.98	0.01	2

The angular diameters were computed by fitting a circular uniform disk to the visibilities from each night.

* Beam combiner codes: C = CLASSIC, CL = CLIMB, M = MIRC.

† CLASSIC/CLIMB H filter: $\lambda = 1.6731 \mu\text{m}$, $\Delta\lambda = 0.2854 \mu\text{m}$; CLASSIC/CLIMB K' filter: $\lambda = 2.1329 \mu\text{m}$, $\Delta\lambda = 0.3489 \mu\text{m}$; MIRC H filter: $\lambda = 1.5\text{--}1.7 \mu\text{m}$ with $\Delta\lambda \approx 0.034 \mu\text{m}$ for each of the eight spectral channels.

‡ Calibrator codes: 1 = HD 190993 ($\theta = 0.161 \pm 0.011 \text{ mas}$; ref. 33), 2 = HD 196740 ($\theta = 0.180 \pm 0.012 \text{ mas}$; ref. 33), 3 = HD 206860 ($\theta = 0.530 \pm 0.015 \text{ mas}$; ref. 34), 4 = HD 190406 ($\theta = 0.584 \pm 0.010 \text{ mas}$; ref. 35), 5 = HD 195810 ($\theta = 0.301 \pm 0.021 \text{ mas}$; ref. 33).

§ Only E2W1W2 and S1S2 were recorded on UT 2013 August 24. Only W1W2, E1E2, S1S2, and E2W2 were recorded on UT 2013 August 26.

Extended Data Table 2 | Uniform ellipse models of Nova Del 2013

UT Date	MJD	t (days)	Comb	θ_{major} (mas)	θ_{minor} (mas)	PA ($^{\circ}$)	χ^2_{ν}
2013 Aug 16	56520.303	2.0	CL	0.631 ± 0.282	0.503 ± 0.243	81.8 ± 31.9	0.14
2013 Aug 17	56521.260	3.0	M	0.676 ± 0.011	0.633 ± 0.007	114.8 ± 10.6	1.30
2013 Aug 19	56523.248	5.0	M	0.936 ± 0.005	0.826 ± 0.004	148.7 ± 2.1	1.53
2013 Aug 21	56525.220	6.9	M	1.291 ± 0.007	1.198 ± 0.005	157.4 ± 3.5	6.95
2013 Aug 23	56527.310	9.0	CL	1.716 ± 0.094	1.535 ± 0.044	138.5 ± 16.1	2.49

Extended Data Table 3 | Infrared magnitudes of Nova Del 2013

MJD	Filt	Magnitude	MJD	Filt	Magnitude	MJD	Filt	Magnitude
56519.179	<i>H</i>	5.12 ± 0.16	56532.170	<i>K'</i>	4.44 ± 0.23	56542.149	<i>K'</i>	5.37 ± 0.16
56519.193	<i>H</i>	5.19 ± 0.17	56532.191	<i>K'</i>	4.53 ± 0.16	56542.158	<i>K'</i>	5.37 ± 0.16
56519.216	<i>H</i>	5.08 ± 0.15	56532.214	<i>H</i>	4.65 ± 0.38	56542.167	<i>K'</i>	5.35 ± 0.15
56519.231	<i>H</i>	5.20 ± 0.14	56533.223	<i>K'</i>	4.65 ± 0.21	56542.176	<i>K'</i>	5.32 ± 0.16
56519.250	<i>K'</i>	4.97 ± 0.17	56533.235	<i>K'</i>	4.60 ± 0.20	56543.125	<i>K'</i>	5.42 ± 0.22
56519.261	<i>K'</i>	5.04 ± 0.17	56534.249	<i>K'</i>	4.74 ± 0.16	56543.135	<i>K'</i>	5.44 ± 0.20
56519.273	<i>K'</i>	4.93 ± 0.16	56534.265	<i>K'</i>	4.78 ± 0.17	56543.150	<i>K'</i>	5.42 ± 0.19
56520.187	<i>H</i>	4.20 ± 0.20	56534.273	<i>K'</i>	4.71 ± 0.16	56543.162	<i>K'</i>	5.36 ± 0.19
56520.218	<i>H</i>	4.24 ± 0.19	56534.404	<i>K'</i>	4.49 ± 0.27	56543.175	<i>K'</i>	5.30 ± 0.18
56520.247	<i>H</i>	4.16 ± 0.17	56537.130	<i>K'</i>	5.08 ± 0.17	56544.123	<i>K'</i>	5.32 ± 0.22
56520.278	<i>H</i>	4.21 ± 0.16	56537.146	<i>K'</i>	4.79 ± 0.18	56544.132	<i>K'</i>	5.44 ± 0.19
56520.303	<i>H</i>	4.15 ± 0.15	56537.157	<i>K'</i>	4.46 ± 0.31	56544.145	<i>K'</i>	5.52 ± 0.19
56520.349	<i>H</i>	4.08 ± 0.16	56539.168	<i>K'</i>	5.10 ± 0.19	56544.156	<i>K'</i>	5.52 ± 0.17
56520.375	<i>H</i>	4.11 ± 0.16	56539.179	<i>K'</i>	5.19 ± 0.22	56544.170	<i>K'</i>	5.42 ± 0.15
56520.412	<i>H</i>	4.07 ± 0.18	56539.190	<i>K'</i>	5.62 ± 0.23	56544.184	<i>K'</i>	5.37 ± 0.17
56520.432	<i>H</i>	4.04 ± 0.19	56540.148	<i>K'</i>	5.25 ± 0.20	56545.128	<i>K'</i>	5.25 ± 0.29
56521.241	<i>H</i>	4.09 ± 0.12	56540.164	<i>K'</i>	5.22 ± 0.18	56545.139	<i>K'</i>	5.46 ± 0.30
56521.263	<i>H</i>	4.05 ± 0.12	56540.179	<i>K'</i>	5.17 ± 0.17	56545.153	<i>K'</i>	6.56 ± 0.54
56522.265	<i>H</i>	3.74 ± 0.15	56540.194	<i>K'</i>	5.14 ± 0.18	56548.144	<i>K'</i>	5.38 ± 0.21
56522.275	<i>H</i>	3.82 ± 0.14	56540.212	<i>K'</i>	5.18 ± 0.16	56548.160	<i>K'</i>	5.26 ± 0.21
56522.284	<i>H</i>	3.75 ± 0.14	56541.143	<i>K'</i>	5.29 ± 0.15	56548.171	<i>K'</i>	5.09 ± 0.22
56522.295	<i>H</i>	3.70 ± 0.14	56541.160	<i>K'</i>	5.31 ± 0.15	56551.119	<i>K'</i>	5.43 ± 0.27
56523.219	<i>H</i>	3.63 ± 0.15	56541.178	<i>K'</i>	5.34 ± 0.15	56551.137	<i>K'</i>	5.49 ± 0.26
56523.248	<i>H</i>	3.46 ± 0.16	56541.195	<i>K'</i>	5.31 ± 0.14	56551.154	<i>K'</i>	5.28 ± 0.22
56524.295	<i>H</i>	3.46 ± 0.16	56541.213	<i>K'</i>	5.22 ± 0.14	56554.118	<i>K'</i>	5.00 ± 0.25
56524.325	<i>H</i>	3.52 ± 0.23	56541.231	<i>K'</i>	5.30 ± 0.12	56554.129	<i>K'</i>	5.08 ± 0.23
56525.189	<i>H</i>	3.79 ± 0.13	56541.247	<i>K'</i>	5.29 ± 0.13	56554.139	<i>K'</i>	5.08 ± 0.22
56525.220	<i>H</i>	3.80 ± 0.14	56541.265	<i>K'</i>	5.34 ± 0.13	56554.150	<i>K'</i>	5.11 ± 0.22
56526.207	<i>K'</i>	3.58 ± 0.15	56541.281	<i>K'</i>	5.30 ± 0.12	56554.159	<i>K'</i>	5.17 ± 0.22
56526.224	<i>K'</i>	3.64 ± 0.13	56541.299	<i>K'</i>	5.22 ± 0.14	56558.102	<i>K'</i>	4.77 ± 0.22
56527.258	<i>K'</i>	3.86 ± 0.13	56541.315	<i>K'</i>	5.10 ± 0.14	56558.112	<i>K'</i>	4.80 ± 0.20
56527.283	<i>K'</i>	3.68 ± 0.13	56541.333	<i>K'</i>	5.41 ± 0.12	56558.121	<i>K'</i>	4.77 ± 0.20
56527.310	<i>K'</i>	3.84 ± 0.16	56541.350	<i>K'</i>	5.19 ± 0.13	56558.131	<i>K'</i>	4.79 ± 0.20
56527.324	<i>K'</i>	3.67 ± 0.16	56541.367	<i>K'</i>	4.86 ± 0.15	56558.139	<i>K'</i>	4.81 ± 0.19
56527.394	<i>H</i>	4.29 ± 0.15	56541.389	<i>K'</i>	5.20 ± 0.14	56561.129	<i>K'</i>	4.87 ± 0.35
56528.241	<i>H</i>	3.91 ± 0.16	56541.409	<i>K'</i>	4.49 ± 0.17	56561.159	<i>K'</i>	4.54 ± 0.28
56530.203	<i>H</i>	4.35 ± 0.12	56542.126	<i>K'</i>	5.38 ± 0.22			
56530.263	<i>H</i>	4.54 ± 0.11	56542.141	<i>K'</i>	5.29 ± 0.17			

Extended Data Table 4 | Effective bandwidth of the CLASSIC/CLIMB observations

UT Date	MJD	t (days)	$\frac{\text{Bandwidth(Nova)}}{\text{Bandwidth(Cal)}}$
2013 Aug 15	56519.231	1.0	0.916 ± 0.295
2013 Aug 16	56520.303	2.0	1.015 ± 0.256
2013 Aug 22	56526.216	7.9	0.970 ± 0.193
2013 Aug 28	56532.191	13.9	1.023 ± 0.134
2013 Aug 29	56533.235	15.0	0.905 ± 0.107
2013 Aug 30	56534.273	16.0	0.879 ± 0.151
2013 Sep 02	56537.146	18.9	0.863 ± 0.104
2013 Sep 04	56539.178	20.9	0.844 ± 0.041
2013 Sep 05	56540.164	21.9	0.844 ± 0.121
2013 Sep 06	56541.280	23.0	0.876 ± 0.208
2013 Sep 07	56542.158	23.9	0.879 ± 0.117
2013 Sep 08	56543.150	24.9	0.817 ± 0.078
2013 Sep 09	56544.156	25.9	0.878 ± 0.086
2013 Sep 10	56545.139	26.9	0.792 ± 0.166
2013 Sep 13	56548.160	29.9	0.836 ± 0.098
2013 Sep 16	56551.137	32.9	0.933 ± 0.194
2013 Sep 19	56554.138	35.9	0.883 ± 0.162
2013 Sep 23	56558.121	39.8	0.902 ± 0.111
2013 Sep 26	56561.159	42.9	1.165 ± 0.492

We present the ratio of the effective bandwidth of the interferometric observations of Nova Del 2013 compared with the calibrator stars.

Experimental realization of the topological Haldane model with ultracold fermions

Gregor Jotzu¹, Michael Messer¹, Rémi Desbuquois¹, Martin Lebrat¹, Thomas Uehlinger¹, Daniel Greif¹ & Tilman Esslinger¹

The Haldane model on a honeycomb lattice is a paradigmatic example of a Hamiltonian featuring topologically distinct phases of matter¹. It describes a mechanism through which a quantum Hall effect can appear as an intrinsic property of a band structure, rather than being caused by an external magnetic field². Although physical implementation has been considered unlikely, the Haldane model has provided the conceptual basis for theoretical and experimental research exploring topological insulators and superconductors^{2–6}. Here we report the experimental realization of the Haldane model and the characterization of its topological band structure, using ultracold fermionic atoms in a periodically modulated optical honeycomb lattice. The Haldane model is based on breaking both time-reversal symmetry and inversion symmetry. To break time-reversal symmetry, we introduce complex next-nearest-neighbour tunnelling terms, which we induce through circular modulation of the lattice position⁷. To break inversion symmetry, we create an energy offset between neighbouring sites⁸. Breaking either of these symmetries opens a gap in the band structure, which we probe using momentum-resolved interband transitions. We explore the resulting Berry curvatures, which characterize the topology of the lowest band, by applying a constant force to the atoms and find orthogonal drifts analogous to a Hall current. The competition between the two broken symmetries gives rise to a transition between topologically distinct regimes. By identifying the vanishing gap at a single Dirac point, we map out this transition line experimentally and quantitatively compare it to calculations using Floquet theory without free parameters. We verify that our approach, which allows us to tune the topological properties dynamically, is suitable even for interacting fermionic systems. Furthermore, we propose a direct extension to realize spin-dependent topological Hamiltonians.

In a honeycomb lattice that is symmetric under time-reversal and inversion, the two lowest bands are connected at two Dirac points. Each broken symmetry leads to a gapped energy spectrum. F. D. M. Haldane realized that the resulting phases are topologically distinct¹: A broken inversion symmetry (IS), caused by an energy offset between the two sublattices, leads to a trivial band insulator at half-filled lattice sites. Time-reversal symmetry (TRS) can be broken by complex next-nearest-neighbour tunnel couplings (Fig. 1a). The corresponding staggered magnetic fluxes sum up to zero in one unit cell, thereby preserving the translation symmetry of the lattice. This gives rise to a topological Chern insulator, where a non-zero Hall conductance appears despite the absence of a net magnetic field^{1,2}. When both symmetries are broken, a topological phase transition connects two regimes with a distinct topological invariant, the Chern number, which changes from 0 to +1 or to −1; see Fig. 1b. There, the gap closes at a single Dirac point. These transitions have attracted great interest because they cannot be described by Landau's theory of phase transitions, owing to the absence of a changing local order parameter⁶.

A crucial experimental challenge for the realization of the Haldane model is the creation of complex next-nearest-neighbour tunnelling. Here we show that this is possible with ultracold atoms in optical lattices periodically modulated in time. Pioneering experiments with bosons showed a renormalization of existing tunnelling amplitudes in one dimension^{9,10}, and were extended to control tunnelling phases^{11,12} and higher-order

tunnelling¹³. In higher dimensions this allowed the study of phase transitions^{14,15}, and topologically trivial staggered fluxes were realized^{16,17}. Furthermore, uniform flux configurations were observed using rotation and laser-assisted tunnelling^{18,19}, although for the latter method, heating seemed to prevent the observation of a flux in some experiments²⁰. In a honeycomb lattice, a rotating force, as proposed by T. Oka and H. Aoki, can induce the required complex tunnelling⁷. Using arrays of coupled waveguides, a classical version of this proposal was used to study topologically protected edge modes in the inversion-symmetric regime²¹. We

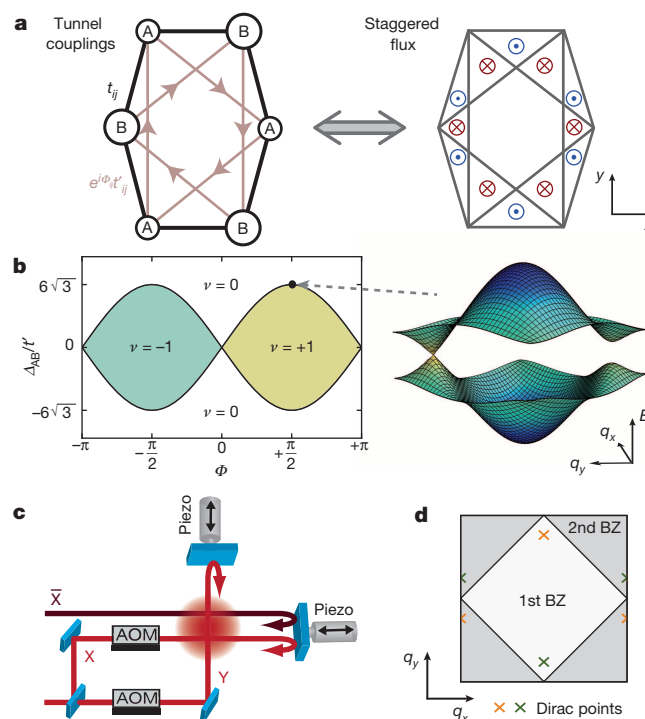


Figure 1 | The Haldane model. **a**, Tight-binding model of the honeycomb lattice realized in the experiment. An energy offset Δ_{AB} between sublattice A and B breaks IS. Nearest-neighbour tunnel couplings t_{ij} have real values, whereas next-nearest-neighbour tunnelling $e^{i\phi_{ij}}t'_{ij}$ carries tunable phases indicated by arrows. i and j indicate the indices of the connected lattice sites. For a detailed discussion of anisotropies and higher-order tunnelling terms, see the Supplementary Information. The corresponding staggered magnetic fluxes (sketched on the right) sum up to zero but break TRS. **b**, Topological regimes of the Haldane model, for isotropic t_{ij} , $t'_{ij} = t'$ and $\Phi_{ij} = \Phi$. The trivial (Chern number $\nu = 0$) and Chern-insulating ($\nu = \pm 1$) regimes are connected by topological transitions (black lines), where the band structure (shown on the right) becomes gapless at a single Dirac point. **c**, Laser beam set-up forming the optical lattice. The laser \bar{X} is frequency-detuned from the other beams. Piezo-electric actuators sinusoidally modulate the retro-reflecting mirrors, with a controllable phase difference ϕ . Acousto-optic modulators (AOMs) ensure the stability of the lattice geometry (Methods). **d**, The resulting Brillouin zones (BZ), featuring two Dirac points in quasi-momentum space \mathbf{q} .

¹Institute for Quantum Electronics, ETH Zurich, 8093 Zurich, Switzerland.

access the full parameter space of the Haldane model using a fermionic quantum gas, by extending the proposal to elliptical modulation of the lattice position and additionally breaking IS through a deformation of the lattice geometry.

The starting point of our experiment is a non-interacting, ultracold gas of 4×10^4 to 6×10^4 fermionic ^{40}K atoms prepared in the lowest band of a honeycomb optical lattice created by several laser beams at wavelength $\lambda = 1,064$ nm, arranged in the x - y plane as depicted in Fig. 1c and detailed in ref. 8. The two lowest bands have a total bandwidth of $h \times 3.9(1)$ kHz (where h is Planck's constant), with a gap of $h \times 5.4(2)$ kHz to the next higher band, and contain two Dirac points at opposite quasi-momenta; see Fig. 1d. After loading the atoms into the honeycomb lattice, we ramp up a sinusoidal modulation of the lattice position along the x and y directions with a final amplitude of $0.087(1)\lambda$, frequency 4.0 kHz and phase difference φ . This gives access to linear ($\varphi = 0^\circ$ or 180°), circular ($\varphi = \pm 90^\circ$) and elliptical trajectories.

The effective Hamiltonian of our system in the phase-modulated honeycomb lattice is computed using analytical and numerical Floquet theory (see the Methods and Supplementary Information for a detailed discussion). It is well described by the Haldane model¹

$$\hat{H} = \sum_{\langle ij \rangle} t_{ij} \hat{c}_i^\dagger \hat{c}_j + \sum_{\langle\langle ij \rangle\rangle} e^{i\Phi_{ij}} t'_{ij} \hat{c}_i^\dagger \hat{c}_j + \Delta_{AB} \sum_{i \in A} \hat{c}_i^\dagger \hat{c}_i \quad (1)$$

where t_{ij} and t'_{ij} are real-valued nearest- and next-nearest-neighbour tunnelling amplitudes, and the latter contain additional phases Φ_{ij} defined along the arrows shown in Fig. 1a. The fermionic creation and annihilation operators are denoted by \hat{c}_i^\dagger and \hat{c}_i . The energy offset $\Delta_{AB} \geq 0$ between sites of the A and B sublattice breaks IS and opens a gap $|\Delta_{AB}|$ (ref. 8). TRS can be broken by changing φ . This controls the imaginary part of the next-nearest-neighbour tunnelling, whereas its real part, as well as t_{ij} and Δ_{AB} , are mostly unaffected (Methods). Breaking only TRS opens an energy gap $|\Delta_T|$ at the Dirac points, given by a sum of the imaginary part of the three next-nearest-neighbour tunnel couplings connecting the same sublattice

$$\Delta_T = - \sum_i w_i t'_i \sin(\Phi_i) = \Delta_T^{\max} \sin(\varphi) \quad (2)$$

with weights w_i of order unity, which depend on the position of the Dirac points in the Brillouin zone. The sum is taken over the different types of next-nearest-neighbour bond, and the origin of the second equality is discussed in the Supplementary Information. Circular modulation ($\varphi = \pm 90^\circ$) leads to a maximum gap ($h \times 88^{+10}_{-34}$ Hz for our parameters), whereas the gap vanishes for linear modulation ($\varphi = 0^\circ, \pm 180^\circ$), where TRS is preserved.

We will first present measurements which confirm that breaking either symmetry is sufficient to open a gap in the band structure. For this, we restrict ourselves to either $\varphi = 0^\circ$ or $\Delta_{AB} = 0$, corresponding to the two axes of the Haldane diagram of Fig. 1b. Subsequently we will present measurements in which we explore the topology of the lowest band in the same parameter regime by probing the Berry curvature. To probe the opening of gaps in the system, we drive Landau-Zener transitions through the Dirac points^{8,22}. Applying a constant force along the x direction by means of a magnetic field gradient causes an energy offset $\Delta E/h = 103.6(1)$ Hz per site, thereby inducing a Bloch oscillation. After one full Bloch cycle the gradient is removed and the fraction of atoms ξ in the second band is determined using a band-mapping procedure (Methods). For broken IS, a gap given by $|\Delta_{AB}|$ opens at both Dirac points. In this case, ξ reaches a maximum at $\Delta_{AB} = 0$, which indicates a vanishing energy gap, and decays symmetrically around this point as expected; see Fig. 2a. In the case of broken TRS (Fig. 2b), a reduction in transfer versus modulation phase is observed. This signals an opening gap, which is found to be largest for circular modulation, as expected from equation (2).

Breaking either IS or TRS gives rise to similar, gapped band structures which remain point-symmetric around quasi-momentum $\mathbf{q} = 0$. However, the energy spectrum itself is not sufficient to reveal the different topology of the band, which is given by the associated eigenstates. These are characterized by a local geometrical property: the Berry curvature $\Omega(\mathbf{q})$ (ref. 6). In \mathbf{q} -space, $\Omega(\mathbf{q})$ is analogous to a magnetic field and corresponds

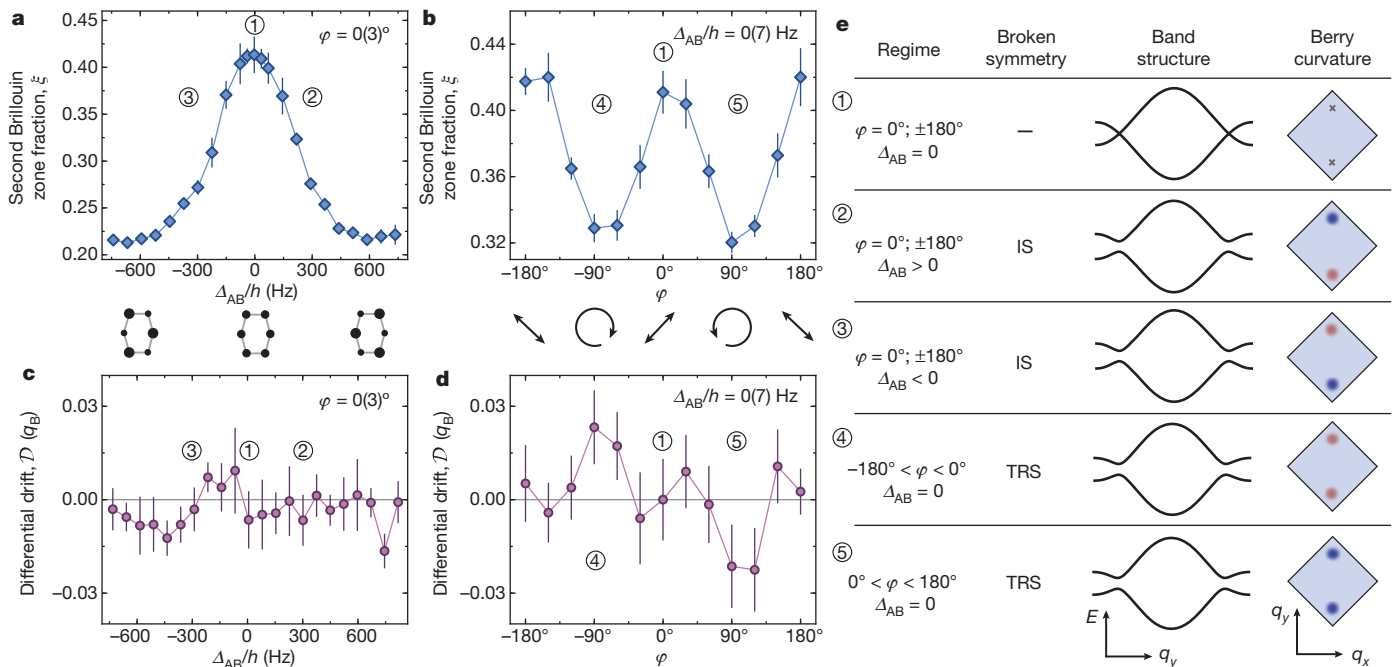


Figure 2 | Probing gaps and Berry curvature. **a, b**, Fraction of atoms in the second band ξ after one Bloch oscillation in the q_x direction. We break either IS (**a**) by introducing a sublattice offset Δ_{AB} or TRS (**b**) via elliptical modulation (see diagrams below). This corresponds to scanning either of the two axes of the Haldane model. A gap opens at both Dirac points, given by $|\Delta_{AB}|$ or $|\Delta_T^{\max} \sin(\varphi)|$, respectively, thereby reducing ξ . **c, d**, Differential drift \mathcal{D} obtained from Bloch oscillations in opposite q_y directions. For broken IS (**c**),

the opposite Berry curvatures at the two Dirac points cancel each other, while for broken TRS (**d**) the system enters the topological regime, where opposite drifts for $\varphi \geq 0$ are expected. Data show mean \pm s.d. of at least 6 (**a–c**) or 21 (**d**) measurements, and the Bloch momentum $q_B = 2\pi/\lambda$. The numbers in parentheses are the standard deviations of the calibrations of the parameters used for those plots. **e**, Sketches illustrating gaps and Berry curvature in different regimes. Red (blue) indicates positive (negative) Berry curvature.

to the geometric phase picked up along an infinitesimal loop. When only IS is broken, the Berry curvature is point-antisymmetric, and its sign inverts for opposite Δ_{AB} ; see Fig. 2e. The spread of $\Omega(\mathbf{q})$ increases with the size of the gap. Its integral over the first Brillouin zone, the Chern number ν , is zero, corresponding to a topologically trivial system. However, with only TRS broken, $\nu = \pm 1$, $\Omega(\mathbf{q})$ is point-symmetric, and its sign changes when reversing the rotation direction of the lattice modulation.

To determine the topology of the lowest band, we move the atoms along the y direction such that their trajectories sample the regions where the Berry curvature is concentrated, and record their final position. As atoms move through regions of \mathbf{q} -space with non-zero curvature, they acquire an orthogonal velocity proportional to the applied force and $\Omega(\mathbf{q})$ (refs 23–26). The underlying harmonic confinement caused by the laser beams in the experiment couples real and momentum space, meaning that a displacement in real space leads to a drift in quasi-momentum. We apply a gradient of $\Delta E/\hbar = 114.6(1)$ Hz per site and measure the centre of mass of the quasi-momentum distribution in the lowest band after one full Bloch cycle. Because the velocity caused by the Berry curvature inverts when inverting the force, we subtract the result for the opposite gradient to obtain the differential drift \mathcal{D} . This quantity is more suitable for distinguishing trivial from non-trivial Berry-curvature distributions than the response to a single gradient (Methods)²⁵. The latter does however provide information about the local Berry curvature and is shown in Extended Data Fig. 2.

When breaking only IS, we observe that \mathcal{D} vanishes and is independent of Δ_{AB} , because the Berry curvature is point-antisymmetric; see Fig. 2c. In contrast, when only TRS is broken, we explore the topological regime of the Haldane model with $\Delta_{AB} = 0$. A differential drift is observed for $\varphi = 90^\circ$, which, as expected, is opposite for $\varphi = -90^\circ$; see Figs 2d and 4c. This is a direct consequence of the Berry curvature being point-symmetric, with its sign given by the rotation direction of the lattice modulation. In fact, here a non-zero \mathcal{D} can only originate from a non-zero integrated Berry curvature (Methods). As the modulation becomes linear, the drift disappears. This is smoothed by the increased transfer to the higher band when the gap becomes smaller, which predominantly affects atoms that would experience the strongest Berry curvature. These observations are qualitatively confirmed by semiclassical simulations shown in Extended Data Fig. 1.

Within the Haldane model, the competition of simultaneously broken TRS and IS is of particular interest, as it features a topological transition between a trivial band insulator and a Chern insulator. In this regime, both the band structure and Berry curvature are no longer point-symmetric and the energy gap G_\pm at the two Dirac points is given by

$$G_\pm = |\Delta_{AB} \pm \Delta_T^{\max} \sin(\varphi)| \quad (3)$$

On the transition lines the system is gapless with one closed and one gapped Dirac point, $G_+ = 0$ or $G_- = 0$. We now discuss measurements in which we extend the parameter regime to allow for the simultaneous breaking of both symmetries.

We map out the transition by measuring the transfer ξ_\pm for each Dirac point separately, see Fig. 3a. ξ_+ (ξ_-) is the fraction of atoms occupying the upper (lower) half of the second Brillouin zone after one Bloch oscillation along the x direction. We observe a difference between ξ_+ and ξ_- , which shows that the band structure is no longer point-symmetric, allowing for the parity anomaly predicted by Haldane¹. When the topology of the band changes, the gap at one of the Dirac point closes. We identify the closing of a gap with the point of maximum measured transfer when varying Δ_{AB} . For $\varphi = 0^\circ$ we find, as expected for preserved TRS, that the maxima of both ξ_+ and ξ_- coincide; see Fig. 3b. The maxima are shifted in opposite directions for $\varphi = 90^\circ$, showing that the minimum gap for each Dirac point occurs at different values of Δ_{AB} . In between these values the system is in the topologically non-trivial regime. We explore the position of each maximum for varying φ and find opposite shifts for negative φ as predicted by equation (3) using no free parameters; see Fig. 3c.

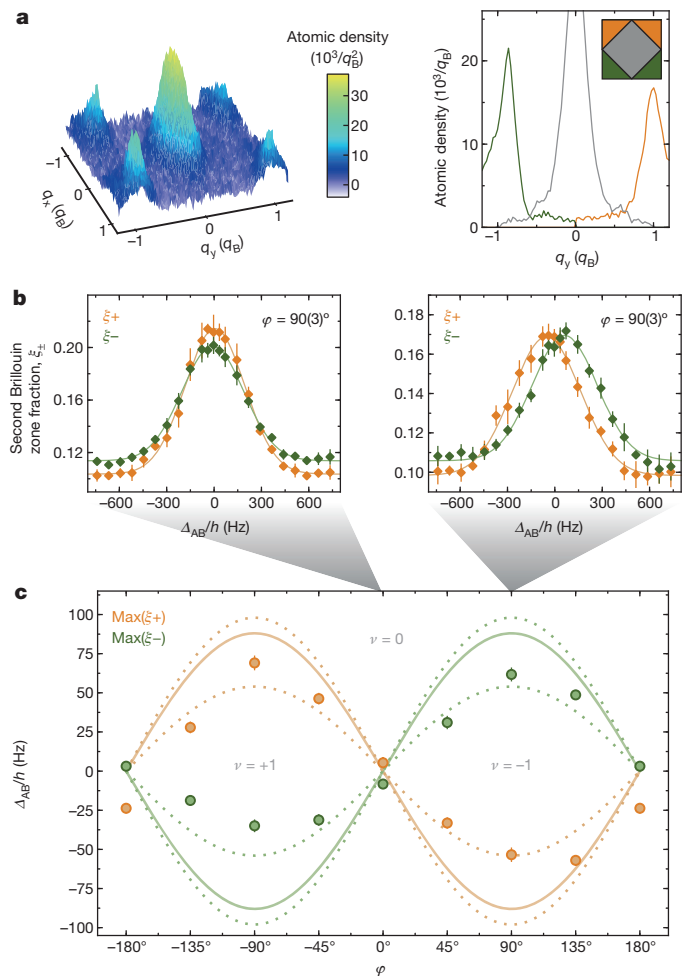


Figure 3 | Mapping out the transition line. **a**, Atomic quasi-momentum distribution (averaged over six runs) after one Bloch oscillation for $\varphi = +90^\circ$, $\Delta_{AB}/h = 292(7)$ Hz. A line sum along q_x shows the atomic density in the first Brillouin zone in grey; atoms transferred at the upper (lower) Dirac point are shown in orange (green) throughout. The fraction of atoms in the second Brillouin zone differs for $q_y \gtrless 0$. This is a direct consequence of simultaneously broken IS and TRS, which allows band structures that are not point-symmetric. **b**, Fractions of atoms ξ_\pm in each half of the second Brillouin zone. For linear modulation (left) the gap vanishes at $\Delta_{AB} = 0$ for both Dirac points, while for circular modulation (right) it vanishes at opposite values of Δ_{AB} . Gaussian fits (solid lines) are used to find the maximum transfer, which signals the topological transition. Data are mean \pm s.d. of at least six measurements. **c**, Solid lines show the theoretically computed topological transitions, without free parameters. Dotted lines represent the uncertainty of the maximum gap $|\Delta_T^{\max}|/h = 88^{+10}_{-34}$ Hz, originating from the uncertainty of the lattice parameters. Data are the points of maximum transfer for each Dirac point, \pm the fitting error, obtained from measurements as in **b** for various φ . Data points for $\varphi = \pm 180^\circ$ correspond to the same measurements. Between the lines, the system is in the topologically non-trivial regime.

In Fig. 4 we show the measured differential drift \mathcal{D} for all topological regimes, allowing for simultaneously broken IS and TRS. Here, we reduce the modulation frequency to 3.75 kHz, where the signal-to-noise ratio of \mathcal{D} is larger, but which is less suited for a quantitative comparison of the transfer ξ because the lattice modulation ramps are expected to be less adiabatic. \mathcal{D} is non-zero only for broken TRS and shows the expected antisymmetry with φ and symmetry with Δ_{AB} . For large Δ_{AB} , deep inside the topologically trivial regime, \mathcal{D} vanishes for all φ . For smaller Δ_{AB} , the differential drift shows precursors of the regimes with non-zero Chern number: non-zero values of \mathcal{D} extend well beyond the transition lines when IS and TRS are both broken. Semiclassical simulations (see Extended Data Fig. 1c) suggest that the main contribution to this effect arises from the transfer to the higher band.

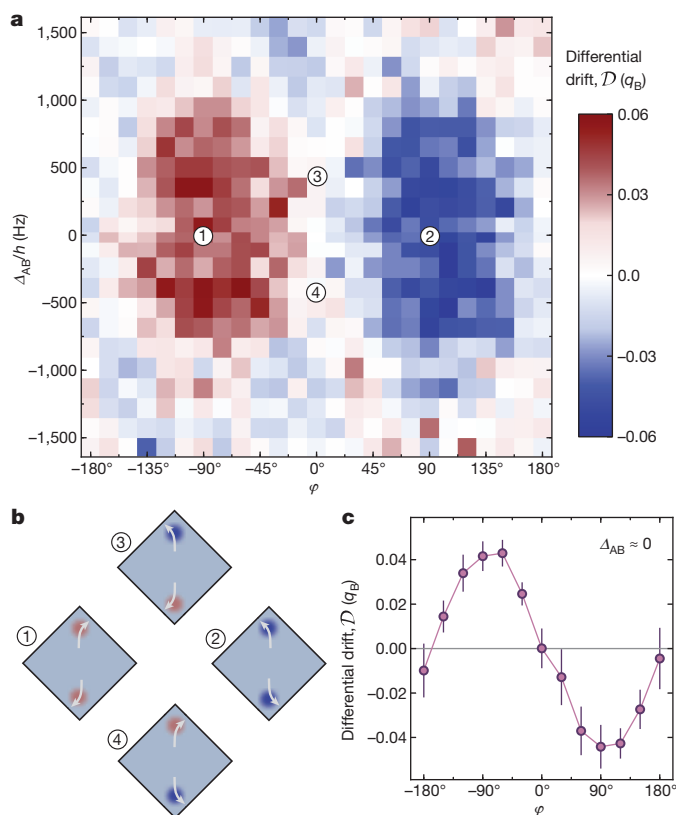


Figure 4 | Drift measurements. **a**, Differential drift \mathcal{D} in quasi-momentum. Each pixel corresponds to at least one pair of measurements, where the modulation frequency was set to 3.75 kHz. Data points for $\varphi = \pm 120^\circ$, $\Delta_{AB}/h = 503(7)$ Hz were not recorded and are interpolated. **b**, All topological regimes are explored and the expected momentum-space drifts caused by the Berry curvature are sketched for selected parameters. **c**, Cut along the $\Delta_{AB}/h = 15(7)$ Hz line. Data show mean \pm s.d. for at least six pairs of measurements.

Extending our work to interacting systems requires sufficiently low heating. We investigate this with a repulsively interacting spin mixture in the honeycomb lattice previously used for studying the fermionic Mott insulator²⁷. We measure the entropy in the Mott insulating regime by loading atoms into the lattice and reversing the loading procedure (see Methods and Extended Data Fig. 3). The entropy increase is only 25% larger than without modulation. This opens up the possibility of studying topological models with interactions²⁸ in a controlled and tunable way. For example, lattice modulation could be used to create topological flat bands, which have been suggested to give rise to interaction-induced fractional Chern insulators^{29,30}. Furthermore, our approach of periodically modulating the system can be directly extended to engineer Hamiltonians with spin-dependent tunnelling amplitudes and phases (Methods). This can be achieved by modulating a magnetic field gradient, which leads to spin-dependent oscillating forces owing to the differential Zeeman shift. For example, TRS topological Hamiltonians, such as the Kane–Mele model³, can be implemented by simultaneously modulating the lattice on one axis and a magnetic field gradient on the other.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 June; accepted 29 September 2014.

1. Haldane, F. D. M. Model for a quantum Hall Effect without Landau levels: condensed-matter realization of the ‘parity’ anomaly. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).

2. Chang, C.-Z. *et al.* Experimental observation of the quantum anomalous Hall effect in a magnetic topological insulator. *Science* **340**, 167–170 (2013).
3. Kane, C. L. & Mele, E. J. Quantum spin Hall effect in graphene. *Phys. Rev. Lett.* **95**, 226801 (2005).
4. König, M. *et al.* Quantum spin Hall insulator state in HgTe quantum wells. *Science* **318**, 766–770 (2007).
5. Hsieh, D. *et al.* A topological Dirac insulator in a quantum spin Hall phase. *Nature* **452**, 970–974 (2008).
6. Hasan, M. Z. & Kane, C. L. Topological insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
7. Oka, T. & Aoki, H. Photovoltaic Hall effect in graphene. *Phys. Rev. B* **79**, 081406 (2009).
8. Tarruell, L., Greif, D., Uehlinger, T., Jotzu, G. & Esslinger, T. Creating, moving and merging Dirac points with a Fermi gas in a tunable honeycomb lattice. *Nature* **483**, 302–305 (2012).
9. Dunlap, D. & Kenkre, V. Dynamic localization of a charged particle moving under the influence of an electric field. *Phys. Rev. B* **34**, 3625–3633 (1986).
10. Lignier, H. *et al.* Dynamical control of matter-wave tunneling in periodic potentials. *Phys. Rev. Lett.* **99**, 220403 (2007).
11. Struck, J. *et al.* Tunable gauge potential for neutral and spinless particles in driven optical lattices. *Phys. Rev. Lett.* **108**, 225304 (2012).
12. Jiménez-García, K. *et al.* Peierls substitution in an engineered lattice potential. *Phys. Rev. Lett.* **108**, 225303 (2012).
13. Parker, C. V., Ha, L.-C. & Chin, C. Direct observation of effective ferromagnetic domains of cold atoms in a shaken optical lattice. *Nature Phys.* **9**, 769–774 (2013).
14. Zenesini, A., Lignier, H., Ciampini, D., Morsch, O. & Arimondo, E. Coherent control of dressed matter waves. *Phys. Rev. Lett.* **102**, 100403 (2009).
15. Struck, J. *et al.* Quantum simulation of frustrated classical magnetism in triangular optical lattices. *Science* **333**, 996–999 (2011).
16. Aidelsburger, M. *et al.* Experimental realization of strong effective magnetic fields in an optical lattice. *Phys. Rev. Lett.* **107**, 255301 (2011).
17. Struck, J. *et al.* Engineering Ising-XY spin-models in a triangular lattice using tunable artificial gauge fields. *Nature Phys.* **9**, 738–743 (2013).
18. Williams, R. A., Al-Assam, S. & Foot, C. J. Observation of vortex nucleation in a rotating two-dimensional lattice of Bose-Einstein condensates. *Phys. Rev. Lett.* **104**, 050404 (2010).
19. Aidelsburger, M. *et al.* Realization of the Hofstadter Hamiltonian with ultracold atoms in optical lattices. *Phys. Rev. Lett.* **111**, 185301 (2013).
20. Miyake, H., Siviloglou, G. A., Kennedy, C. J., Burton, W. C. & Ketterle, W. Realizing the Harper Hamiltonian with laser-assisted tunneling in optical lattices. *Phys. Rev. Lett.* **111**, 185302 (2013).
21. Rechtsman, M. C. *et al.* Photonic Floquet topological insulators. *Nature* **496**, 196–200 (2013).
22. Lim, L.-K., Fuchs, J.-N. & Montambaux, G. Bloch-Zener oscillations across a merging transition of Dirac points. *Phys. Rev. Lett.* **108**, 175303 (2012).
23. Chang, M.-C. & Niu, Q. Berry phase, hyperorbits, and the Hofstadter spectrum. *Phys. Rev. Lett.* **75**, 1348–1351 (1995).
24. Dudarev, A. M., Diener, R. B., Carusotto, I. & Niu, Q. Spin-orbit coupling and Berry phase with ultracold atoms in 2D optical lattices. *Phys. Rev. Lett.* **92**, 153005 (2004).
25. Price, H. M. & Cooper, N. R. Mapping the Berry curvature from semiclassical dynamics in optical lattices. *Phys. Rev. A* **85**, 033620 (2012).
26. Dauphin, A. & Goldman, N. Extracting the Chern number from the dynamics of a Fermi gas: implementing a quantum Hall bar for cold atoms. *Phys. Rev. Lett.* **111**, 135302 (2013).
27. Uehlinger, T. *et al.* Artificial graphene with tunable interactions. *Phys. Rev. Lett.* **111**, 185307 (2013).
28. Varney, C. N., Sun, K., Rigol, M. & Galiatski, V. Interaction effects and quantum phase transitions in topological insulators. *Phys. Rev. B* **82**, 115125 (2010).
29. Neupert, T., Santos, L., Chamon, C. & Mudry, C. Fractional quantum Hall states at zero magnetic field. *Phys. Rev. Lett.* **106**, 236804 (2011).
30. Grushin, A. G., Gómez-León, A. & Neupert, T. Floquet fractional Chern insulators. *Phys. Rev. Lett.* **112**, 156801 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Aoki for drawing our attention to the relevance of their proposal for optical lattices and N. Cooper, S. Huber, L. Tarruell, L. Wang and A. Zenesini for discussions. We acknowledge the SNF, the NCCR-QSIT and the SQMS (ERC advanced grant) for funding.

Author Contributions The data were measured by G.J., M.M., R.D. and D.G. and analysed by G.J., M.M., R.D., T.U. and D.G. The theoretical framework was developed by G.J. and M.L. All work was supervised by T.E. All authors contributed to planning the experiment, discussions and the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.E. (esslinger@phys.ethz.ch).

METHODS

Spin-polarized Fermi gas. After sympathetic cooling with ^{87}Rb in a magnetic trap, 1×10^6 fermionic ^{40}K atoms are transferred into an optical dipole trap operating at a wavelength of 826 nm. A balanced spin mixture of the $|F, m_F\rangle = |9/2, -9/2\rangle$ and $|9/2, -7/2\rangle$ Zeeman states, where F denotes the hyperfine manifold and m_F the magnetic sub-level, is evaporatively cooled at a magnetic field of 197.6(1) G, in the vicinity of the Feshbach resonance located at 202.1 G. We obtain typical temperatures of $0.2T_F$, where T_F is the Fermi temperature. The field is subsequently reduced and a magnetic field gradient is used to selectively remove the $|9/2, -7/2\rangle$ component, while levitating the atoms in the $|9/2, -9/2\rangle$ state against gravity.

Honeycomb optical lattice. This polarized Fermi gas is loaded into the optical lattice within 200 ms using an S-shaped intensity ramp, and the dipole trap is subsequently turned off in 100 ms. The lattice potential is given by⁸

$$V(x, y, z) = -V_X \cos^2(k_{\text{lat}}x + \theta/2) - V_X \cos^2(k_{\text{lat}}x) \\ - V_Y \cos^2(k_{\text{lat}}y) - 2\alpha\sqrt{V_X V_Y} \cos(k_{\text{lat}}x) \cos(k_{\text{lat}}y) \cos\varphi_L \quad (4) \\ - V_Z \cos^2(k_{\text{lat}}z)$$

where $V_{X,Y,Z}$ are the single-beam lattice depths and $k_{\text{lat}} = 2\pi/\lambda$. Gravity points along the negative y -direction. To control the energy offset \mathcal{A}_{AB} , we vary θ around π by changing the frequency detuning δ between the X and the Y (which has the same frequency as Y) beams using an acousto-optic modulator; see Fig. 1c. The point where $\mathcal{A}_{AB} = 0$ is determined from transfer measurements in a static lattice with an accuracy of 7 Hz. The phase φ_{lat} is stabilized to $0.0(3)^\circ$ using a heterodyne interferometer⁸, and the visibility of the interference pattern is $\alpha = 0.81(1)$. We minimize the intensity imbalance between the incoming and reflected lattice beams in the x - y plane such that the remaining imbalance between left and right vertical tunnelling is less than 0.3%, as determined from Raman-Nath diffraction on a ^{87}Rb Bose-Einstein condensate. The final lattice depths are set to $V_{X,Y,Z} = [5.0(3), 0.45(2), 2.3(1), 0] E_R$, where the recoil energy $E_R = \hbar^2/2m\lambda^2$, and m denotes the mass of ^{40}K . Using a Wannier function calculation²⁷, we extract the corresponding tight-binding parameters $t_{0,1,2,3}/\hbar = [-746(81), -527(17), -527(17), -126(7)]$ Hz, for the horizontal, the left and right vertical nearest-neighbour tunnelling links, and the horizontal link across the honeycomb cell, respectively (see Supplementary Information). This results in a bandwidth of $W/\hbar = 3.9(1)$ kHz. The amplitudes for the next-nearest-neighbour tunnel couplings are $t'_{1,2,3}/\hbar = [14, 14, 61]$ Hz and do not affect the gaps and the topological transition line. The nearest-neighbour tunnelling is renormalized in the modulated lattice, which decreases the effective bandwidth to $W_{\text{eff}}/\hbar = 3.3(1)$ kHz.

All experiments are carried out in the presence of a weak underlying harmonic confinement with trapping frequencies $\omega_{x,y,z}/2\pi = [14.4(6), 30.2(1), 29.3(3)]$ Hz, which originates from the Gaussian intensity profiles of the red-detuned lattice beams. The lattice depths are calibrated using Raman-Nath diffraction on a ^{87}Rb Bose-Einstein condensate. To determine α , we drive quasi-momentum resolved inter-band transitions for a spin-polarized Fermi gas loaded into a chequerboard lattice by periodically modulating the lattice depth, and we measure the band gap at $\mathbf{q} = 0$. **Modulation of the optical lattice.** The two mirrors used for retro-reflecting the lattice beams are mounted on piezo-electric actuators, which allow for a controlled phase shift of the reflected beams with respect to the incoming lattice beams. To fix the geometry of the lattice, the relative phase φ_{lat} of the two orthogonal retro-reflected beams X and Y is actively stabilized to $\varphi_{\text{lat}} = 0^\circ$. To maintain this phase relation during modulation, the phase of the respective incoming beams is modulated at the same frequency as the piezo-electric actuators using acousto-optical modulators. In addition, this phase modulation provides a direct calibration of the amplitude and relative phase of the mirror displacement. The calibration is confirmed by measuring both the reduction of tunnelling¹⁰ and the effective atomic mass around $\mathbf{q} = 0$ in a modulated simple cubic lattice.

The modulation is turned on as follows: the atoms are loaded into a lattice with 30% larger single-beam lattice depths than the final values used for the actual measurements. This suppresses resonant transfer of atoms to higher bands. The modulation amplitude is then linearly increased within 20 ms to reach a normalized drive of $K_0 = 0.7778$, where $K_0 = \pi^2(A/\lambda)(\hbar\omega/E_R)$, with A the amplitude of the motion, $\omega/2\pi$ the modulation frequency and $\hbar = \hbar/2\pi$. The time-dependence of the lattice position $\mathbf{r}_{\text{lat}}(t)$ is then given by

$$\mathbf{r}_{\text{lat}} = -A(\cos(\omega t)\mathbf{e}_x + \cos(\omega t - \varphi)\mathbf{e}_y) \quad (5)$$

where \mathbf{e}_x and \mathbf{e}_y denote the real-space unit vectors along the x and y directions. The phase φ is set with an accuracy of 3° . When using different modulation frequencies, we keep K_0 constant. The lattice depths are then finally reduced to their final values within 10 ms. We have independently confirmed that the modulation frequency of $\omega/2\pi = 4$ kHz exceeds the combined static bandwidth of the two lowest bands using lattice phase-modulation spectroscopy.

We have verified that our experimental findings are not affected by the global phase of the lattice modulation or changes in the total modulation time smaller than a modulation period. This means that a time-independent effective Hamiltonian can safely be used. Additionally, we have varied the timescale on which the modulation is ramped on and off to confirm that these ramps are sufficiently slow and the measurements are not affected by switch-on or switch-off effects.

Detection. After one Bloch cycle lasting 9.85 ms (8.72 ms) for x (y) oscillations, the lattice modulation amplitude is linearly lowered to zero within 2 ms. The quasi-momentum distribution of the atoms is then probed using a band-mapping technique, where all lattice beams are ramped down within 500 μs , which is much shorter than the timescales of the harmonic trap, meaning that the original \mathbf{q} -space distribution is conserved⁸. An absorption image of the atomic distribution is then recorded after 15 ms of ballistic expansion. The fraction of atoms per band is determined by integrating the atomic density in the corresponding Brillouin zone in the absorption image. The size of the Brillouin zone is independently calibrated by using Bloch oscillations of a non-interacting Fermi gas in a one-dimensional lattice. Owing to the residual non-adiabaticity of the lattice ramps, 16% (14%) of the atoms are detected in the second band and 21% (8%) in even higher bands after loading the lattice and before the Bloch cycle including (not including) the linear ramps of the modulation.

For the drift measurements, the displacement of the atoms with respect to the position before the Bloch oscillation is obtained by calculating the centre of mass within the first Brillouin zone. The differential drift is then calculated as the difference of the recorded drift for oscillations along the positive and negative q_y direction. We already observe individual drifts when only breaking IS, which constitutes a measurement of a local Berry curvature (see below and Extended Data Fig. 2). The drift depends strongly on the size of the energy gap, which supports our explanation for the precursors of the topological phase in Fig. 4.

Effective band-structure calculations. The effective Hamiltonian \hat{H}_{eff} is given by a logarithm of the time-evolution operator for one modulation period. Numerically, we discretize time and \mathbf{q} -space, choosing the grid such that a higher resolution does not further change the results. Analytically, we use a Magnus expansion in $1/\omega$. The energy spectrum, Berry curvature and Chern number are computed from \hat{H}_{eff} using the usual methods employed for static Hamiltonians. By comparing with numerical results and computing higher-order terms we show that the Magnus expansion can be truncated at first order in $1/\omega$ for our parameters, and \hat{H}_{eff} then takes on the form of the Haldane Hamiltonian (equation (1)).

There, \mathcal{A}_{AB} is not affected by the modulation compared to the static lattice and t_{ij} are renormalized by a factor of about 0.85 with variations of ± 0.03 , depending on φ and the orientation of the tunnelling. The induced imaginary next-nearest-neighbour tunnelling takes on values of up to $\hbar \times 18$ Hz for the diagonal links and $\hbar \times 5$ Hz for the vertical links. Its value depends only on the modulation amplitude (scaling as a product of two second-order Bessel functions), the frequency (scaling as $1/\omega$), and the product of two nearest-neighbour tunnel couplings which correspond to this next-nearest-neighbour tunnelling. Static real next-nearest neighbour tunnel couplings do not affect other terms in \hat{H}_{eff} and are not required to open a gap, meaning that our approach works equally well in deep optical lattices. The weights w_i determining the gap in equation (3) are 3.5 and 2.1 for the vertical and diagonal tunnel couplings respectively, with variations of about ± 0.1 as a function of φ . Detailed derivations, formulae for all terms in the effective Hamiltonian, and comparisons between numerical and analytical approaches can be found in the Supplementary Information.

Semi-classical calculations. We use a semi-classical approximation to simulate the orthogonal drifts observed for Bloch oscillations along the y direction in the experiment, using the same lattice parameters as in Fig. 2. We denote the energy and Berry curvature of the lowest band of the analytical effective Hamiltonian with $\epsilon(q_x, q_y)$ and $\Omega(q_x, q_y)$. The external accelerating force is given by $F_y = \pm 2\Delta E/\lambda$, where $\Delta E/\hbar = 114.6$ Hz is the energy offset per site. The equations of motion then read (omitting the z -direction, which de-couples)

$$\dot{x} = \frac{1}{\hbar} \partial_{q_x} \epsilon(q_x, q_y) - \dot{q}_y \cdot \Omega(q_x, q_y) \quad (6)$$

$$\dot{y} = \frac{1}{\hbar} \partial_{q_y} \epsilon(q_x, q_y) + \dot{q}_x \cdot \Omega(q_x, q_y) \quad (7)$$

$$\hbar \dot{q}_x = -\partial_x V_{\text{trap}}(x, y, z) \quad (8)$$

$$\hbar \dot{q}_y = F_y - \partial_y V_{\text{trap}}(x, y, z) \quad (9)$$

The effect of the harmonic trap is taken into account with $V_{\text{trap}}(x, y, z) = 0.5m(\omega_x^2 x^2 + \omega_y^2 y^2 + \omega_z^2 z^2)$.

As the underlying band structure possesses several symmetries, the possible values of the differential drift \mathcal{D} are strongly constrained by the topology of the lowest band. When only TRS (IS) is broken, the band structure is point-symmetric, $\epsilon(q_x, q_y) = \epsilon(-q_x, -q_y)$. Additionally, the Berry curvature is point-symmetric (point-antisymmetric), so the lowest band is topologically non-trivial (trivial). If the system is also reflection-symmetric, $\epsilon(q_x, q_y) = \epsilon(-q_x, q_y)$, the equation of motion for \dot{q}_x remains unchanged when inverting the direction of the force in the topologically trivial case. Thus, a differential drift can only appear if the lowest band is topologically non-trivial. In the experiment, these symmetries are strictly present when $\varphi = \pm 90^\circ$. The modulation weakly breaks reflection symmetry otherwise (see Supplementary Information) but it can be restored by considering the average of $\mathcal{D}(\varphi)$ and $\mathcal{D}(\pi - \varphi)$. For both cases the experimental data of Figs 2d and 4c shows that the lowest band is topologically non-trivial.

For the numerical simulations, we compute 4×10^4 trajectories, starting from a zero-temperature fermionic phase-space distribution (taking into account the z -direction) resembling the measured initial momentum-space distribution, and then determine the \mathbf{q} -space centre-of-mass position after one Bloch cycle. For a rough estimate of the effects of Landau-Zener transfers to the second band, we record the minimum bandgap experienced by each trajectory and exclude trajectories below a chosen cut-off value. This approach will not capture the complex quantum-mechanical dynamics of the real transfer process, but may serve to indicate in which direction the measured drift curves will be deformed.

Extended Data Fig. 1a shows that no differential drift is expected when only IS is broken, as observed in the experiment. In particular, even though reflection symmetry is weakly broken in the system as stated above, its effect remains smaller than the numerical error on \mathcal{D} . The simulations furthermore show that for each individual applied force a drift should appear, as observed (see below and Extended Data Fig. 2a). When only TRS is broken, a differential drift that changes sign with the modulation phase φ is computed, which is smaller but comparable to the measured values. The sudden change of \mathcal{D} around $\varphi = 0^\circ$ is smoothed when taking into account transfer to higher bands. In that region the gap at the Dirac points and the spread of the Berry curvature is very small, meaning that the atoms which would contribute most to the drift are likely to be transferred to the higher band. A similar effect appears in Extended Data Fig. 1c, where both IS and TRS are broken for circular modulation. Without transfer, our simulations predict a sudden change in \mathcal{D} at the topological transition. When taking transfer into account, the Dirac point with the smaller gap contributes less, so the drifts observed in the topological regime extend beyond the transition line, as measured in the experiment (see Fig. 4a). This transfer could be reduced by applying weaker gradients to the atomic cloud, which would, however, require removing the harmonic trapping potential along the y direction.

The underlying harmonic trap in the x direction is also of particular importance, as it is responsible for transforming displacements in real space into momentum-space drifts. As seen in Extended Data Fig. 1d, the differential drift vanishes when ω_x does, increases rapidly close to the experimental value of $\omega_x = 2\pi \times 14.4(6)$ Hz and shows oscillatory behaviour when the timescale of dipole oscillations becomes comparable to the Bloch oscillation period.

Drift measurements for opposite forces. Now we discuss the individual drift observed after one full Bloch cycle supplementing Fig. 2c and d and Fig. 4 in the main text. Our measurement technique probes the Berry curvature of the lowest band by moving atoms in the q_y direction past the gapped Dirac points, where the Berry curvature is localized, and relies on the coupling of real and momentum space. Note that the width of the fermionic cloud is sufficiently large that by performing a Bloch oscillation we sample the entire Brillouin zone. As described in the main text, the applied uniform force causes an orthogonal velocity in real space when atoms are in the region of the Berry curvature. The resulting change in real space position induces a transverse force arising from the underlying harmonic confinement in the opposite direction to the direction of the displacement. As a result, we observe a drift in quasi-momentum, which is amplified by the negative effective mass at each Dirac point arising from the negative curvature of the band structure.

In the topologically trivial case, when only IS is broken, for each individual gradient we observe an equal drift along q_x when scanning the sublattice offset Δ_{AB} (see Extended Data Fig. 2a). This constitutes a measure of the local Berry curvature, because the integrated Berry curvature is zero. We measure a drift which increases with increasing gap and changes sign with Δ_{AB} . The data shows that the observed drift predominantly arises from the first Dirac point that is passed. This effect can be explained in the following way: The Berry curvature of the first Dirac point already leads to an orthogonal drift in quasi-momentum. When successively reaching the second Dirac point the shifted part of the cloud then does not experience the same Berry curvature distribution. As expected, opposite oscillation directions give rise to the same drift, since not only the direction of the force changes but also the sign of the Berry curvature corresponding to the first Dirac point on the trajectory (see

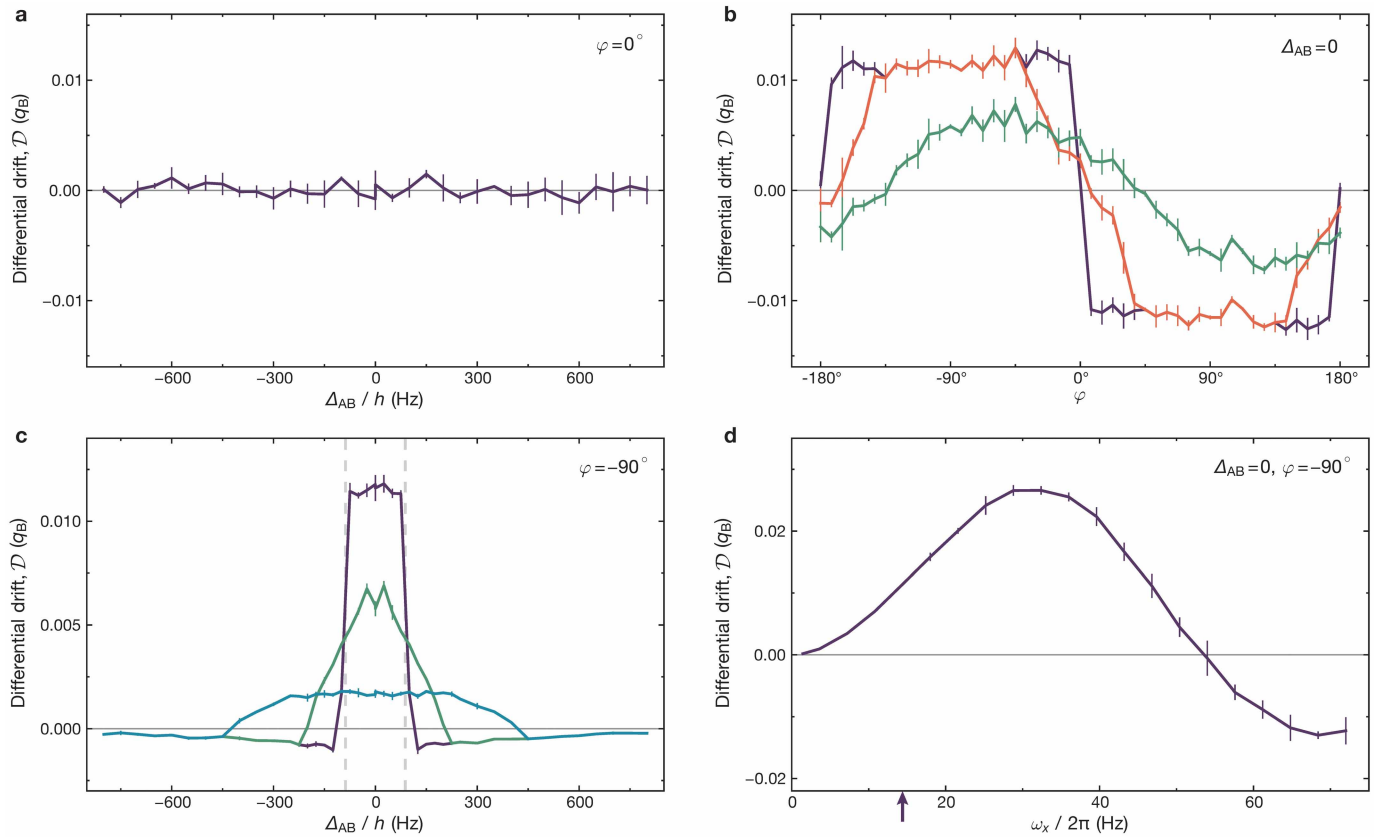
Extended Data Fig. 2). Subtracting the drifts of opposite gradients gives the differential drift \mathcal{D} , as shown in the main text. This probes the net contribution of both Dirac points. Only for the largest Δ_{AB} does the measured drift decrease again, indicating an increasing spread of the Berry-curvature distributions at each Dirac point, which then start to overlap and gradually cancel each other. We also observe these drifts in a static lattice with $\Delta_{AB} \neq 0$.

In contrast, when only TRS is broken, we probe the topologically non-trivial regime. Extended Data Fig. 2b shows opposite drifts along q_x for each of the oscillation directions. In this case both successively passed Dirac points cause a drift in the same direction since the Berry curvature is point-symmetric. Therefore, changing the sign of the applied gradient leads to a drift in the opposite q_x direction. As expected, the drift changes sign for the opposite modulation phase difference φ , directly revealing the changing sign of the Berry curvature distribution. Here too, a larger gap (when φ is closer to $\pm 90^\circ$ or the modulation frequency is lower) leads to a larger drift. The combination of the dependence of the drift on the size of the gap and the predominance of the first Dirac point on the trajectory may be the cause of the precursors seen in Fig. 4a for simultaneously broken IS and TRS.

Heating of an interacting Fermi gas. We have also investigated the increase in entropy in the modulated honeycomb lattice in comparison to the static case by loading a repulsively interacting Fermi gas into the lattice and reversing the loading procedure³¹. We prepare about $2.0(2) \times 10^5$ atoms in a balanced spin mixture of the $|F, m_F\rangle = |9/2, -9/2\rangle$ and $|9/2, -5/2\rangle$ Zeeman states. The atoms are then loaded into a lattice with final depths $V_{\bar{x}, \bar{y}, \bar{z}} = [14.0(4), 0.79(2), 6.45(20), 7.0(2)]E_R$ within 200 ms. This corresponds to a system consisting of coupled isotropic honeycomb layers with a nearest-neighbour tunnelling $t/h = 172(20)$ Hz, as used in previous work²⁷. The two lowest bands have a total bandwidth of $h \times 1.0$ kHz, with a gap of $h \times 14$ kHz to the next higher band (excluding the third direction). After turning on the lattice modulation (using the same K_0 as in the main text, which opens a gap of about $h \times 44$ Hz at the Dirac points), we reverse the loading procedure and measure the final temperature of the sample. This is compared to the case where the lattice is not modulated. From the difference in temperature before loading the lattice and after the procedure, the corresponding entropy increase can be determined. We measure the entropy increase for different interaction strengths and modulation frequencies. In the Mott-insulating regime with $U/h = 4.18(2)$ kHz ($U/5t = 4.9(6)$) and for a frequency of $\omega/2\pi = 1.08$ kHz (which is the same in proportion to the bandwidth as in the measurements of the main text) we find an entropy increase that is 25% larger when modulating the lattice compared to the situation without modulation; see Extended Data Fig. 3a. In the crossover regime at $U/h = 2.19(5)$ kHz ($U/5t = 2.5(3)$) we find the same final entropy. This now corresponds to a 40% increase, which possibly originates from the creation of low-energy charge excitations for these parameters. In the measurements versus frequency, we cover the doublon excitation peak in the insulating phase, whose frequency is given by $U/h = 4.18(2)$ kHz. We have furthermore measured the additional heating induced by holding the atoms in the modulated lattice for longer times; see Extended Data Fig. 3b. For all parameters we find a linear increase of entropy with time. For timescales relevant for studying dynamics, as in the measurements of the main text, this contribution is much smaller than the one associated with the modulation ramp. We find that the heating induced by the modulation does not dominate the final temperature, thus demonstrating that the scheme is well suited for studying many-body states in topological lattices.

Proposal for creating spin-dependent Hamiltonians. To realize spin-dependent Hamiltonians—for example, the Kane–Mele model—we propose to use an oscillating magnetic gradient in order to apply an oscillating spin-dependent force on the atoms. For the $|9/2, -9/2\rangle$ state of ^{40}K , maximum gradients of about 10 G cm^{-1} would be required to achieve the same K_0 as in the main text when modulating with a frequency of 1 kHz. Owing to the Zeeman effect, another spin component or atomic species would experience a different force. If the magnetic moment of the second component is chosen to be opposite, the Kane–Mele model is simply realized by replacing one oscillating mirror in the experiment with this oscillating gradient. The two spin components would then experience clockwise or anti-clockwise modulated forces, respectively, and therefore the two spin bands would have opposite Chern numbers. In general, a combination of an oscillating mirror and magnetic gradient can be used to create the desired average and differential force for other combinations of magnetic moments. This approach can be extended to bosonic atoms or Bose–Fermi mixtures and is also suitable for creating other types of spin-dependent tunnelling, for example, situations where one species is pinned to the lattice and the other remains itinerant.

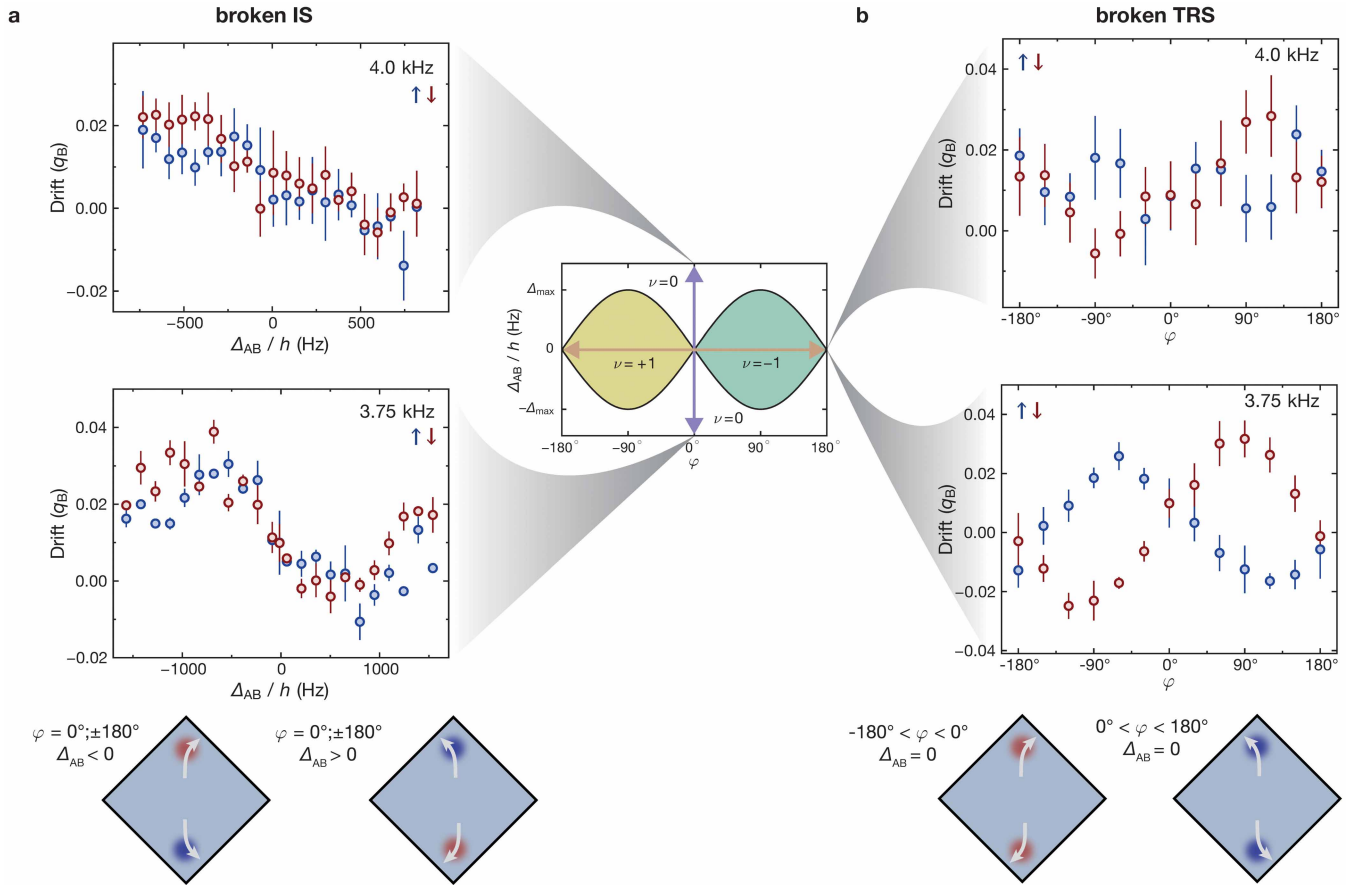
31. Greif, D., Tarruell, L., Uehlinger, T., Jördens, R. & Esslinger, T. Probing nearest-neighbor correlations of ultracold fermions in an optical lattice. *Phys. Rev. Lett.* **106**, 145302 (2011).



Extended Data Figure 1 | Semi-classical simulations of the atomic motion.

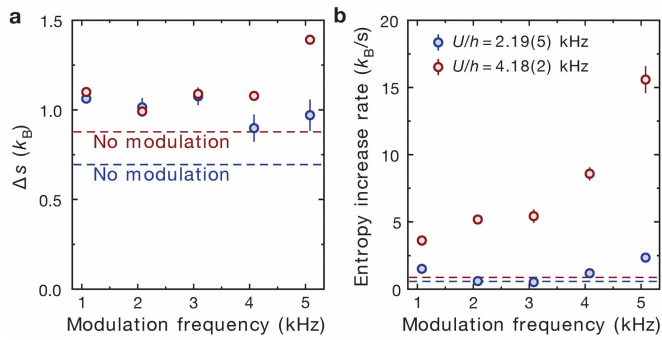
The experiments shown in the main text in Fig. 2c and d are simulated using the semi-classical equations of motion (equations (6)–(9)). The band structure and the Berry curvature are those of the effective Haldane Hamiltonian (equation (1)). The atomic ensemble is modelled by a zero-temperature Fermi distribution. Data are mean \pm s.d. of three simulations containing 4×10^4 trajectories. The differential drift \mathcal{D} is computed when breaking either IS (a) or TRS (b). The former shows no differential drift, in agreement with the experimental data of Fig. 2c. For the latter, the different curves take into account the transfer to the higher band by excluding trajectories passing through regions where the bandgap lies below a certain threshold. If this transfer is not taken into account (purple line), the differential drift varies sharply around $\varphi = 0^\circ$ where the Chern number changes. However, as the threshold is raised to

$0.5\Delta E$ (red line) and ΔE (green line) where $\Delta E/h = 114.6$ Hz is the energy offset per site driving the Bloch oscillation, this sharp feature progressively smoothens and qualitatively reproduces the experimental measurements. c, When TRS is maximally broken ($\varphi = 90^\circ$) and Δ_{AB} varies, the transfer is also responsible for the differential drift extending beyond the topological phase. Without any transfer (purple line), the differential drift changes sharply around the topological phase transition (vertical dashed line), while it extends significantly in the topologically trivial phase when the threshold is set at ΔE (green line) or $3\Delta E$ (blue line). d, Influence of the transverse trapping frequency $\omega_x/2\pi$. The frequency used in the experiment is indicated by a purple arrow. For much larger frequencies the differential drift can vanish, as the transverse oscillation time becomes comparable to the Bloch period.



Extended Data Figure 2 | Drift measurement for broken IS and TRS. The measured drift used to obtain the differential drift \mathcal{D} in Figs 2c, d and 4 of the main text is individually shown for positive and negative forces in the q_y direction. Data for positive (negative) force is shown in blue (red). The central plot, showing the Haldane phase diagram, indicates the region which is scanned when breaking either IS (purple arrow) or TRS (brown arrow) in our system. **a**, We break IS by introducing a sublattice offset and show measurements with modulation frequency of 4.0 kHz and 3.75 kHz. Although the opposite Berry curvatures at the two Dirac points sum up to zero within the first Brillouin zone, we clearly see a drift depending on the size of Δ_{AB} . Data show mean \pm s.d. of at least 6 (4.0 kHz) or 2 (3.75 kHz) measurements. **b**, By changing the modulation phase difference φ we break TRS and the system

enters the topologically non-trivial regime. Drift data for positive (negative) force is shown in blue (red) for a modulation frequency of 4.0 kHz and 3.75 kHz. Data show mean \pm s.d. of at least 21 (4.0 kHz) or 6 (3.75 kHz) measurements. Schematics below show the expected orthogonal drifts caused by driving the atoms through the Berry curvature distribution. Red (blue) indicates positive (negative) Berry curvature. If only IS is broken (**a**) the Berry curvature distribution is point-antisymmetric and changes sign when changing the sign of the sublattice offset. For opposite forces this leads to the same direction of the drift, as indicated by the white arrows. If only TRS is broken (**b**) the Berry curvature distribution at each Dirac point has the same sign, which is changed when reversing the rotation direction. In this case the opposite forces lead to opposite directions of the drift.



Extended Data Figure 3 | Heating of a repulsively interacting Fermi gas.

a, Entropy increase associated with loading into the modulated lattice and reversing the loading procedure. **b**, Entropy increase rate in the modulated lattice for long holding times. The modulation frequency $\omega = 2\pi \times 1,080$ Hz opens a gap of $h \times 44$ Hz in the non-interacting band-structure. This value, in units of the tunnelling, is similar to the measurements of the main text. The dashed lines show the measured heating in a lattice without modulation with identical interaction strengths.

Observation of topological transitions in interacting quantum circuits

P. Roushan^{1†*}, C. Neill^{1*}, Yu Chen^{1†*}, M. Kolodrubetz², C. Quintana¹, N. Leung¹, M. Fang¹, R. Barends^{1†}, B. Campbell¹, Z. Chen¹, B. Chiaro¹, A. Dunsworth¹, E. Jeffrey^{1†}, J. Kelly¹, A. Megrant¹, J. Mutus^{1†}, P. J. J. O'Malley¹, D. Sank^{1†}, A. Vainsencher¹, J. Wenner¹, T. White¹, A. Polkovnikov², A. N. Cleland¹ & J. M. Martinis^{1,3}

Topology, with its abstract mathematical constructs, often manifests itself in physics and has a pivotal role in our understanding of natural phenomena. Notably, the discovery of topological phases in condensed-matter systems has changed the modern conception of phases of matter^{1–5}. The global nature of topological ordering, however, makes direct experimental probing an outstanding challenge. Present experimental tools are mainly indirect and, as a result, are inadequate for studying the topology of physical systems at a fundamental level. Here we employ the exquisite control afforded by state-of-the-art superconducting quantum circuits to investigate topological properties of various quantum systems. The essence of our approach is to infer geometric curvature by measuring the deflection of quantum trajectories in the curved space of the Hamiltonian⁶. Topological properties are then revealed by integrating the curvature over closed surfaces, a quantum analogue of the Gauss–Bonnet theorem. We benchmark our technique by investigating basic topological concepts of the historically important Haldane model⁷ after mapping the momentum space of this condensed-matter model to the parameter space of a single-qubit Hamiltonian. In addition to constructing the topological phase diagram, we are able to visualize the microscopic spin texture of the associated states and their evolution across a topological phase transition. Going beyond non-interacting systems, we demonstrate the power of our method by studying topology in an interacting quantum system. This required a new qubit architecture^{8,9} that allows for simultaneous control over every term in a two-qubit Hamiltonian. By exploring the parameter space of this Hamiltonian, we discover the emergence of an interaction-induced topological phase. Our work establishes a powerful, generalizable experimental platform to study topological phenomena in quantum systems.

Since the first observations of topological ordering in quantum Hall systems in the 1980s^{1,2}, experimental studies of topological phases have mainly been performed with a limited number of primarily indirect measurement techniques. For instance, transport measurements are the predominant tool used to study the quantum Hall effect, where interpretations¹⁰ are required to infer topological properties from the measurements. Consequently, topological studies in quantum systems where transport measurements cannot be carried out have remained elusive.

In principle, topological properties can be explored in any quantum system where the Hamiltonian can be written in terms of a set of parameters. Topological phases are characterized by topological invariants, such as the first Chern number $\mathcal{C}\mathcal{H}$, whose discrete jumps indicate transitions between different topologically ordered phases^{11,12}. For a quantum system, $\mathcal{C}\mathcal{H}$ is defined as the integral over a closed manifold \mathcal{S} in the parameter space of the Hamiltonian as

$$\mathcal{C}\mathcal{H} \equiv \frac{1}{2\pi} \oint_{\mathcal{S}} \mathbf{B} \cdot d\mathbf{S} \quad (1)$$

where \mathbf{B} is the vector form of the Berry curvature¹³. As illustrated in Fig. 1 and shown in Supplementary Information, \mathbf{B} can be viewed as an

effective magnetic field with points of ground-state degeneracy acting as its sources, that is, magnetic monopoles¹⁴. Using Gauss's law for the Berry curvature (magnetic field), $\mathcal{C}\mathcal{H}$ simply counts the number of degenerate energy eigenvalues (magnetic monopoles) enclosed by the parameter manifold \mathcal{S} . $\mathcal{C}\mathcal{H}$, which is invariant under perturbations to the shape of \mathcal{S} , is a topological number that reflects a property of the manifold of states as a whole and not a local property of parameter space.

In previous works, topological properties of highly symmetric quantum systems have been measured^{15–17}. However, since these earlier studies relied on interference, these methods are not readily generalizable. To circumvent this, Gritsev *et al.*⁶ proposed a general method to directly measure the local Berry curvature. The underlying physics of their idea is that motion in a curved space will be deflected from a straight trajectory; in other words, curvature reveals itself as an effective force, analogous to a charged particle moving in a magnetic field experiencing the well-known Lorentz force. Similarly, Gritsev *et al.* showed that in a region of the parameter space with Berry curvature \mathbf{B} , if we 'move' a quantum system by changing a parameter of its Hamiltonian with rate $|\dot{\mathbf{v}}|$, then the state of the system feels a force \mathbf{F} given by

$$\mathbf{F} \propto \mathbf{v} \times \mathbf{B} + \mathbf{O}(\mathbf{v}^2), \quad (2)$$

where \mathbf{O} indicates higher-order terms. This force leads to deviations of the trajectory from the adiabatic path, which can be detected through measurements of the observables of the system (Fig. 1). As long as the

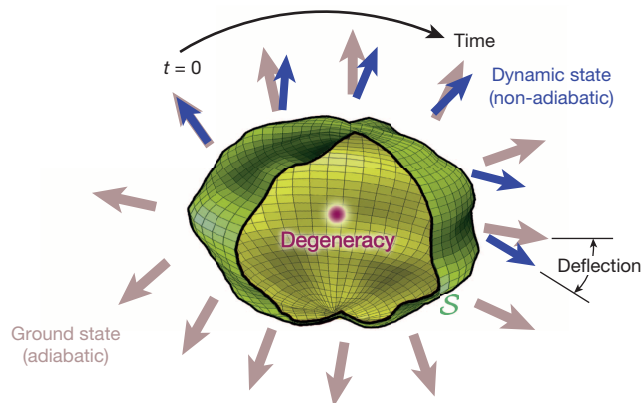


Figure 1 | Dynamical measurement of Berry curvature and $\mathcal{C}\mathcal{H}$. In this schematic drawing, brown arrows represent the ground states (adiabatic limit) for given points on a closed manifold \mathcal{S} (green enclosure, interrupted by an opening for the sake of illustration) in the Hamiltonian's parameter space, and the blue arrows are the measured states during a non-adiabatic passage. According to equation (2) in the main text, the Berry curvature \mathbf{B} can be calculated from the deflection from adiabaticity. Integrating \mathbf{B} over \mathcal{S} gives the Chern number $\mathcal{C}\mathcal{H}$, which corresponds to the total number of degeneracy points (such as the brown point) enclosed.

¹Department of Physics, University of California, Santa Barbara, California 93106-9530, USA. ²Department of Physics, Boston University, Boston, Massachusetts 02215, USA. ³Google Inc., Santa Barbara, California 93117, USA. [†]Present address: Google Inc., Santa Barbara, California 93117, USA.

*These authors contributed equally to this work.

ramping of parameters is done slowly, but not necessarily adiabatically, the deviation is directly proportional to \mathbf{B} . Since the adiabatic limit is generally hard to achieve, this relation has the important advantage of needing only a moderately slow change of state and only requires that the linear term dominates the response.

Direct measurement of \mathbf{B} provides an alternative means to study topological phases that differs significantly from conventional approaches. In condensed-matter systems an instantaneous realization of the entire phase space manifold, such as the Fermi surface, is required. In our approach, the local curvature of the space is dynamically ‘sensed’ and topological invariants, such as $\mathcal{C}\mathcal{H}$, are inferred by integrating these measurements. Implementing this dynamical procedure requires the ability to continuously change the system Hamiltonian. In fully controllable quantum systems, where this can be achieved, this method provides a powerful means to probe topological properties.

To elucidate this dynamical method, we demonstrate a basic implementation in quantum circuits with superconducting qubits^{18–20}. The quantum state of a single qubit²¹ is equivalent to a spin-1/2 particle in a magnetic field. Its Hamiltonian in the rotating frame can be written as

$$\mathcal{H}_S = -\frac{\hbar}{2} \mathbf{H} \cdot \boldsymbol{\sigma}, \quad (3)$$

where $\boldsymbol{\sigma} = (\sigma^x, \sigma^y, \sigma^z)$ are the Pauli matrices, and $\mathbf{H} = (H_x, H_y, H_z)$ is analogous to a control magnetic field. Full control over the parameters of this Hamiltonian is achieved by microwave pulses that control H_x and H_y , and an applied flux through the qubit’s SQUID (superconducting quantum interference device) loop which controls H_z . We measure $\mathcal{C}\mathcal{H}$ for spherical ground-state manifolds in \mathbf{H} parameter space (Fig. 2). We use θ and ϕ as spherical coordinates and consider the parameter trajectory that starts at the north pole at $t = 0$ and ramps along the $\phi = 0$ meridian ($H_y = 0$) with constant velocity $v_\theta = d\theta/dt$ until it reaches the south pole at final time $t = T_f$. To realize motion on a spherical manifold, the control sequences of H_z and H_x are chosen such that the control magnitude $|\mathbf{H}| = H_r$ is constant. In the adiabatic limit, the wavefunction would remain in the instantaneous ground state of \mathcal{H}_S , that is, the $\phi = 0$ meridian on the Bloch sphere. For non-adiabatic ramps, instead, a deviation from the meridian is observed, as shown in Fig. 2b. Here the Bloch vector is measured at each point in time by interrupting the ramp and performing state tomography. Note that this deviation is not due to noise, but rather is the expected non-adiabatic response due to local Berry curvature. For this trajectory, the force \mathbf{F} takes the form $f_\phi = \frac{\hbar}{2} H_r \langle \sigma^y \rangle \sin \theta$, and integrating over the resulting deflection (shaded light red in Fig. 2b) gives $\mathcal{C}\mathcal{H} = 1 \pm 0.05$. Given the symmetry of this Hamiltonian, a line integral is sufficient for measuring the surface integral of $\mathcal{C}\mathcal{H}$ (refs 22, 23). A value of unity is expected, as the qubit ground state has a degeneracy at $\mathbf{H} = 0$, corresponding to a single monopole enclosed by the parameter sphere S . We demonstrate the robustness of $\mathcal{C}\mathcal{H}$ by deforming the surface manifold S (see Supplementary Information).

The generality of our approach allows us to connect our measurements to certain condensed-matter systems and their core topological features, such as topological phase transitions and the geometric winding of state vectors. This can be done by establishing a mapping from the real or momentum space of the model condensed-matter system to the parameter space of the controllable quantum circuit. We choose what is perhaps the simplest theoretical model of topological behaviour, the Haldane model⁷, to benchmark our approach. This model serves as a foundation for other topological insulator models^{3–5}. To show that the quantum Hall effect could be achieved without a global magnetic field, Haldane introduced a non-interacting Hamiltonian⁷ given by

$$\mathcal{H}_G(k_x, k_y) = \hbar v_F (k_x \sigma^x + k_y \sigma^y) + (m_0 - m_t) \sigma^z \quad (4)$$

where v_F is the Fermi velocity, k_x, k_y are the (momentum-space) coordinates, m_0 is the effective mass, and m_t corresponds to a second-neighbour hopping (tunnelling) in a local magnetic field. The key prediction of the Haldane model is that if $m_0/m_t > 1$ the system is in a trivial insulating

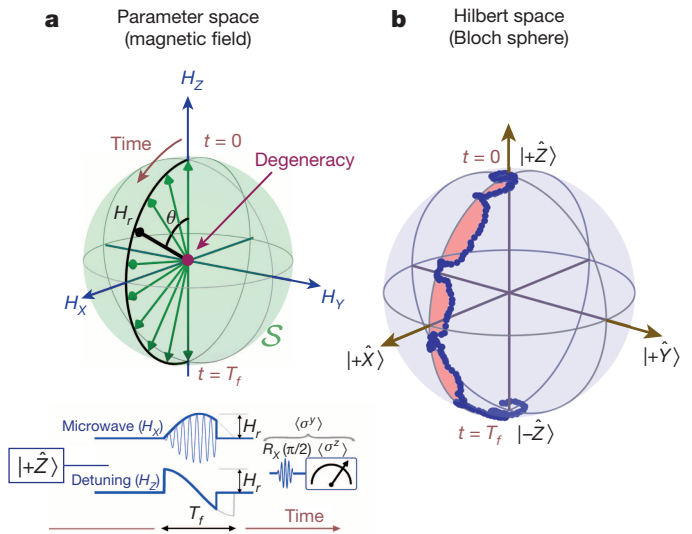


Figure 2 | Dynamical measurement of $\mathcal{C}\mathcal{H}$. **a**, A simultaneous microwave pulse $H_x(t) = H_r \sin(\pi t/T_f)$ and detuning pulse $H_z(t) = H_r \cos(\pi t/T_f)$ are applied to construct a parameter-space trajectory. The pulse sequence results in a parameter-space motion along the $\phi = 0$ meridian ($H_y = 0$ plane) on S . The lower panel shows the pulse sequence applied to the system from preparation to measurement. **b**, The state of the qubit during this ramp ($H_r/2\pi = 10$ MHz and $T_f = 600$ ns) is determined using tomography, and shown (blue dots) on the surface of the Bloch sphere. The light red shading shows the deflection integrated over, yielding $\mathcal{C}\mathcal{H} = 1 \pm 0.05$.

phase, and otherwise in a topological phase. Using a confocal mapping (see Supplementary Information), one can recast equation (4) into the single-qubit Hamiltonian of equation (3). If we consider spherical manifolds S of radius H_r displaced from the origin in the z direction by H_0 , then H_0/H_r in the qubit system plays the same role as m_0/m_t in the Haldane model.

In Fig. 3a we plot the results of this measurement, showing $\mathcal{C}\mathcal{H}$ as a function of H_r and H_0 , which shows plateaux at values 0 and 1 separated by a phase transition boundary line at $H_r = H_0$. This transition can be easily understood: when $H_0 < H_r$, the degeneracy at $\mathbf{H} = 0$ lies within S giving $\mathcal{C}\mathcal{H} = 1$, whereas for $H_0 > H_r$, it lies outside S giving $\mathcal{C}\mathcal{H} = 0$.

In the Haldane model, the topological and the trivial phase each has its own signature spin texture in momentum space. Microscopic structure of these phases can be revealed by the conventional adiabatic method. We again consider spherical surfaces S and adiabatically ramp the control parameters to their final values on S . The resulting Bloch vectors are then tomographically measured. With a confocal mapping (see Supplementary Information), S can be mapped to the first Brillouin zone (FBZ) of the honeycomb lattice. Therefore, the adiabatically measured ground-state vectors on S can be depicted in the FBZ. Figure 3b, c shows the results for two manifolds with $H_0/H_r = 1.2$ and 0, corresponding to trivial and topological phases, respectively. By following the orientation of the state-vector along any path starting at \mathbf{K} and moving to \mathbf{K}' (corners of the FBZ) and back to \mathbf{K} , one can see that in the topological case the state vector wraps around and makes one full rotation, while in the trivial case it only tilts away from vertical and then returns, without completing a rotation. These spin texture maps can also be used to extract local Berry curvature. As shown in Fig. 3c, the resulting $\mathcal{C}\mathcal{H}$ from this adiabatic method shows good agreement with the dynamical method of measurement.

Some of the most fascinating topological phenomena in quantum systems emerge in the presence of interaction. Compared to non-interacting systems, interactions impose a greater experimental challenge to studying topological properties. Nevertheless, the method outlined here stands out in its ability to provide insight into the topology of such systems. To illustrate this, we consider the most basic unit of interaction and measure $\mathcal{C}\mathcal{H}$ in a coupled two-qubit system. One major experimental challenge

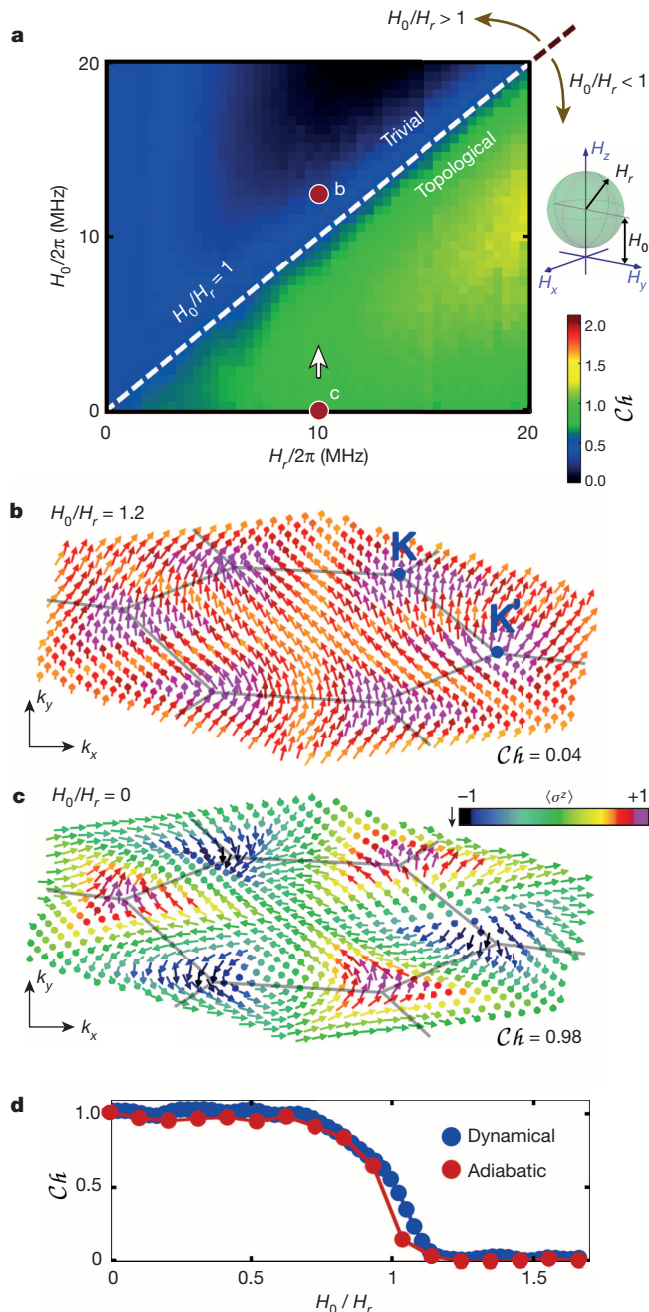


Figure 3 | Dynamic measurement of the topological phase diagram and adiabatic visualization of phases. **a**, Dynamical determination of the phase diagram. First $\langle \sigma^z \rangle$ was measured during ramps similar to those in Fig. 2a, and then \mathcal{Ch} was calculated. The dashed line is the expected phase boundary at $H_0 = H_r$. The ramp speed was $T_f = 1,000$ ns. Red dots show locations of the states whose properties are shown in panels **b** and **c**. **b**, **c**, With adiabatic state preparation, the state of the qubit was prepared and measured over a grid on the surface of the parameter sphere and then mapped to the hexagonal momentum-space plane. The ground states are presented as Bloch vectors, whose colours indicate their $\langle \sigma^z \rangle$ values (see key). $H_0/H_r = 1.2$ for **b** and $H_0/H_r = 0$ for **c**. The grey lines show the FBZ of the honeycomb lattice and high symmetry points K and K' are marked. Each adiabatic sequence took $T_f = 1,000$ ns. **d**, The measured \mathcal{Ch} from the adiabatic and dynamical (white arrow in **a**) methods are plotted versus H_0/H_r .

here is the need for full control over the entire parameter space of the Hamiltonian. Here we achieve this by using a new design for our superconducting qubits, which includes the ability to continuously vary the inter-qubit coupling strength g (refs 8, 9). The Hamiltonian of this system in a frame rotating with the qubits is given by

$$\mathcal{H}_{2Q} = -\frac{\hbar}{2} [H_0 \sigma_1^z + \mathbf{H}_1 \cdot \boldsymbol{\sigma}_1 + \mathbf{H}_2 \cdot \boldsymbol{\sigma}_2 - g(\sigma_1^x \sigma_2^x + \sigma_1^y \sigma_2^y)] \quad (5)$$

where 1 and 2 refer to qubit 1 (Q1) and qubit 2 (Q2), respectively, and the biasing field H_0 is now only applied to Q1. In this design, we can access all regions of the seven-dimensional parameter space of this Hamiltonian.

We explore spherical manifolds with fixed $(H_0, |\mathbf{H}_1|, |\mathbf{H}_2|, g)$, analogous to the single-qubit experiment. We perform experiments where both $\mathbf{H}_1 = \mathbf{H}_2 = \mathbf{H}_r$ are ramped simultaneously with magnitude $|\mathbf{H}_r| = H_r$ (Supplementary Information). The measured \mathcal{Ch} is shown in Fig. 4a, c for three distinct cuts through this parameter space.

We begin in Fig. 4a by exploring the simplest case, $g = 0$, where the qubits behave independently and the physics can be understood using the single qubit results. Since only Q1 is subject to H_0 , its \mathcal{Ch} changes by 1 through the transition $H_0 = H_r$. In contrast, in the absence of a biasing field, \mathcal{Ch} of Q2 remains equal to 1. As the qubits are independent, the \mathcal{Ch} of the system is simply the summation of the individual \mathcal{Ch} , leading to two phases with $\mathcal{Ch} = 1$ and $\mathcal{Ch} = 2$.

With the non-interacting limit of our system understood, we now focus on the effects of interaction by examining regions of parameter space where $g \neq 0$. Considering manifolds with constant $g/2\pi = 4$ MHz, we observe a new phase with $\mathcal{Ch} = 0$ (blue) when $H_r \lesssim g$, as shown in Fig. 4a. To gain more insight into this new phase, we continuously vary g and examine the evolution of the $\mathcal{Ch} = 0$ region. As shown in Fig. 4c, this phase appears when $g \approx H_r$ and monotonically expands as g increases. These observations and the fact that this phase is absent when $g = 0$ indicate that the $\mathcal{Ch} = 0$ phase is indeed driven by interaction.

In certain limits, the three phases could approximately be characterized by the dominance of the global field ($\mathcal{Ch} = 2$), of local fields (disorder; $\mathcal{Ch} = 1$), and of interaction ($\mathcal{Ch} = 0$). Interestingly, they also show some signature entanglement entropies (see Supplementary Information). The linear entropy of the states, averaged over the manifold, qualitatively hints towards a similar phase diagram in certain regions, where the phase with lowest \mathcal{Ch} appears when the highest entanglement allowed in the system has been reached. However, since \mathcal{Ch} is a global property, information about it cannot be deduced from the nature of any single ground state. The interplay of fields and interactions provides hints to anticipate the various topological phases in this system, but are incapable of capturing the entire underlying physics that leads to quantized \mathcal{Ch} values. Therefore, by reflecting topological attributes of the system, \mathcal{Ch} remains distinct and irreplaceable.

As the analogy with Gauss's law suggests, a concrete understanding of the phases can be obtained by considering how the singularities of the system move in the parameter space. Given the relatively small size of the Hilbert space, analytic solutions can be obtained and the phase diagram can be predicted by calculating when points with degenerate ground states cross the spherical manifold. The points of ground-state degeneracy are located on the z axis of \mathbf{H}_r space. In Fig. 4c, the small diagrams at right (A, B and C) correspond to the dots labelled A, B and C on the main panel, where g is small. In this limit, H_0 affects only one qubit, and increasing it moves only one monopole past the surface (C). For D, E and F where instead H_0 is small, increasing g further the monopole separation, eventually moving both monopoles outside the surface (F). The results of a full analytical study are plotted in Fig. 4b, which shows three distinct regions and their phase boundaries. There is a direct 0-to-2 transition when $H_0 = 0$, but at finite values the system first goes through the green $\mathcal{Ch} = 1$ region. This latter behaviour is seen in Fig. 4c. The dashed lines in Fig. 4a, c are from this analytic solution, which uses no free parameters, and are in good agreement with the measurements. The deviations are mainly systematic errors, due to crosstalk between simultaneous control pulses.

The generality of our method is aligned with Feynman's original idea of quantum simulation²⁴, where a controllable quantum system is used to investigate otherwise inaccessible quantum phenomena. In the absence of other experimental approaches, the full controllability of our superconducting circuits will provide a unique platform for the exploration

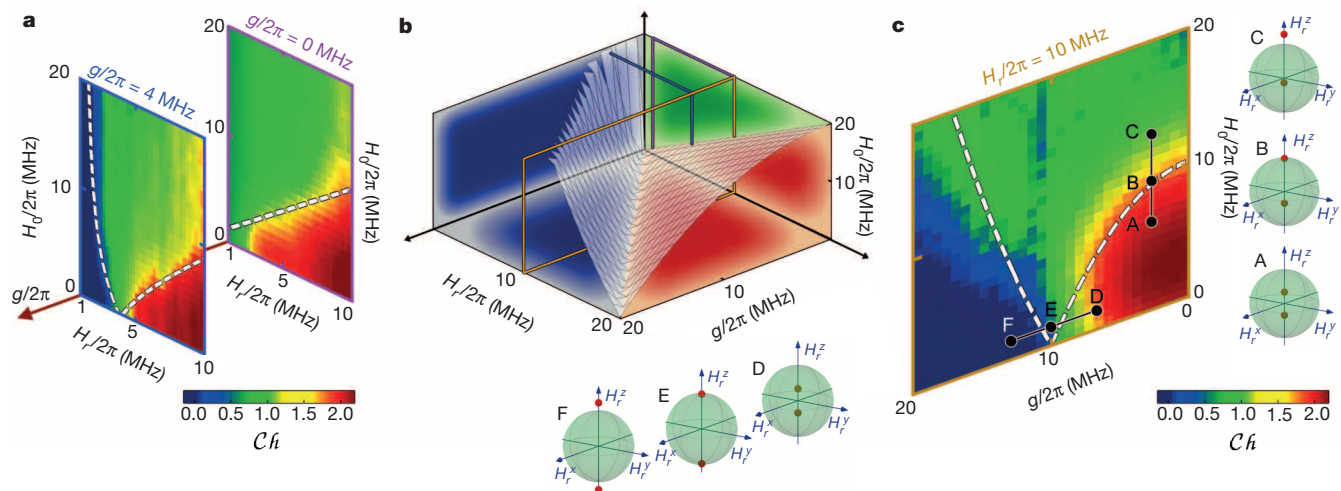


Figure 4 | Topological phase diagram of an interacting system.

a–c, Measured (**a**, **c**) and calculated (**b**) phase diagrams of main text equation (5). In **a**, C_h was measured for two fixed $g/2\pi$ values of 4 MHz (left) and 0 MHz (right); in **c**, C_h was measured for fixed $H_y/2\pi = 10$ MHz. Dashed lines are topological transitions calculated analytically. **b**, The analytically

of topological phases of more complex quantum systems, perhaps most notably interacting spin systems where tantalizing evidence for fractionalization has been found⁶. Larger spin systems have smaller energy gaps, and longer ramps will be needed to remain close to the ground-state manifold. A path forward is therefore to improve coherence in multi-qubit systems, research which is currently underway.

Received 24 June; accepted 23 September 2014.

1. Klitzing, K. v., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980).
2. Tsui, D. C., Stormer, H. L. & Gossard, A. C. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
3. Bernevig, B. A., Hughes, T. L. & Zhang, S.-C. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* **314**, 1757–1761 (2006).
4. Hasan, M. Z. & Kane, C. L. Colloquium: topological insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
5. Moore, J. E. The birth of topological insulators. *Nature* **464**, 194–198 (2010).
6. Gritsev, V. & Polkovnikov, A. Dynamical quantum Hall effect in the parameter space. *Proc. Natl Acad. Sci. USA* **109**, 6457–6462 (2012).
7. Haldane, F. D. M. Model for a quantum Hall effect without Landau levels: condensed-matter realization of the “parity anomaly”. *Phys. Rev. Lett.* **61**, 2015–2018 (1988).
8. Geller, M. *et al.* Tunable coupler for superconducting Xmon qubits: perturbative nonlinear model. Preprint at <http://arXiv.org/abs/1405.1915> (2014).
9. Chen, Y. *et al.* Qubit architecture with high coherence and fast tunable coupling. Preprint at <http://arXiv.org/abs/1402.7367> (2014).
10. Thouless, D. J., Kohmoto, M., Nightingale, M. P. & den Nijs, M. Quantized Hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982).
11. Wen, X.-G. *Quantum Field Theory of Many-body Systems* (Oxford Univ. Press, 2004).
12. Bernevig, B. A. & Hughes, T. L. *Topological Insulators and Topological Superconductors* (Princeton Univ. Press, 2013).
13. Berry, M. V. Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. Lond. A* **392**, 45–57 (1984).
14. Wilczek, F. & Shapere, A. *Geometric Phases in Physics* (World Scientific, 1989).

calculated phase diagram showing three distinct C_h volumes and the separatrix plane. The phase diagram cuts in **a** and **c** are indicated by coloured slices. In **c**, the six spheres show the positions of the monopoles in H -space at the indicated locations A–F in the phase diagram of **c**, with radius equalling the fixed $H_y/2\pi = 10$ MHz.

15. Neeley, M. *et al.* Emulation of a quantum spin with a superconducting phase qubit. *Science* **325**, 722–725 (2009).
16. Leek, P. J. *et al.* Observation of Berry’s phase in a solid-state qubit. *Science* **318**, 1889–1892 (2007).
17. Atala, M. *et al.* Direct measurement of the Zak phase in topological Bloch bands. *Nature Phys.* **9**, 795–800 (2013).
18. Mariani, M. *et al.* Implementing the quantum von Neumann architecture with superconducting circuits. *Science* **334**, 61–65 (2011).
19. Houck, A. A., Türeci, H. E. & Koch, J. On-chip quantum simulation with superconducting circuits. *Nature Phys.* **8**, 292–299 (2012).
20. Buluta, I. & Nori, F. Quantum simulators. *Science* **326**, 108–111 (2009).
21. Barends, R. *et al.* Coherent Josephson qubit suitable for scalable quantum integrated circuits. *Phys. Rev. Lett.* **111**, 080502 (2013).
22. Schroer, M. D. *et al.* Measuring a topological transition in an artificial spin-1/2 system. *Phys. Rev. Lett.* **113**, 050402 (2014).
23. Xu, C., Poudel, A. & Vavilov, M. G. Nonadiabatic dynamics of a slowly driven dissipative two-level system. *Phys. Rev. A* **89**, 052102 (2014).
24. Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge discussions with J. Moore, C. Nayak, M. Niu, A. Rahmani, T. Souza, M. Vavilov, D. Weld and A. Yazdani. This work was supported by the NSF (grants DMR-0907039 and DMR-1029764), the AFOSR (FA9550-10-1-0110), and the ODNI, IARPA, through ARO grant W911NF-10-1-0334. Devices were made at the UCSB Nanofab Facility, part of the NSF-funded NNIN, and the NanoStructures Cleanroom Facility.

Author Contributions P.R., C.N. and Y.C. designed and fabricated the sample, performed the experiment, analysed the data, and with M.K. and C.Q. co-wrote the manuscript and Supplementary Information. M.K. and A.P. provided theoretical assistance. All members of the UCSB team contributed to the experimental set-up. All authors contributed to the manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M.M. (martinis@physics.ucsb.edu).

Interfacial mode coupling as the origin of the enhancement of T_c in FeSe films on SrTiO_3

J. J. Lee^{1,2*}, F. T. Schmitt^{1*}, R. G. Moore^{1*}, S. Johnston^{3,4,5}, Y.-T. Cui¹, W. Li¹, M. Yi^{1,2}, Z. K. Liu^{1,2}, M. Hashimoto⁶, Y. Zhang^{1,7}, D. H. Lu⁶, T. P. Devereaux¹, D.-H. Lee^{8,9} & Z.-X. Shen^{1,2}

Films of iron selenide (FeSe) one unit cell thick grown on strontium titanate (SrTiO_3 or STO) substrates have recently shown^{1–4} superconducting energy gaps opening at temperatures close to the boiling point of liquid nitrogen (77 kelvin), which is a record for the iron-based superconductors. The gap opening temperature usually sets the superconducting transition temperature T_c , as the gap signals the formation of Cooper pairs, the bound electron states responsible for superconductivity. To understand why Cooper pairs form at such high temperatures, we examine the role of the SrTiO_3 substrate. Here we report high-resolution angle-resolved photoemission spectroscopy results that reveal an unexpected characteristic of the single-unit-cell FeSe/ SrTiO_3 system: shake-off bands suggesting the presence of bosonic modes, most probably oxygen optical phonons in SrTiO_3 (refs 5–7), which couple to the FeSe electrons with only a small momentum transfer. Such interfacial coupling assists superconductivity in most channels, including those mediated by spin fluctuations^{8–14}. Our calculations suggest that this coupling is responsible for raising the superconducting gap opening temperature in single-unit-cell FeSe/ SrTiO_3 .

The dramatic enhancement of the superconducting transition temperature T_c in FeSe, from 8 K in bulk¹⁵ to nearly 70 K when grown as a single-unit-cell (1UC) layer on SrTiO_3 , has generated research interest because it suggests that Cooper pairing can be strengthened in non-bulk systems. We begin by studying the electronic structure of the 1UC FeSe/ SrTiO_3 film. Growth details are discussed in the Supplementary Information and Extended Data Fig. 1. The electronic structure is plotted in Fig. 1a–f and is consistent with previous reports^{2–4}. The Fermi surface of the 1UC film consists of electron-like (concave up, labelled A) pockets centred around the Brillouin zone corner (M-point) with a band bottom 60 meV below the Fermi energy (E_F). Further analysis reveals that there are two nearly degenerate electron bands at M (see Supplementary Information). Below the bottom of A, one hole-like (concave down, labelled B) band is clearly resolved. The zone centre (Γ) consists of another hole-like band (labelled D) with a top located 80 meV below E_F . The temperature evolution of the band structure is shown in Fig. 2a–f and shows typical superconducting gap behaviour, with the electron band bending away from the Fermi energy (backbending) at the Fermi momentum k_F . The symmetrized energy distribution curves (EDCs) shown in Fig. 2c are fitted using a phenomenological model (ref. 16) to reveal a 13-meV gap plotted in Fig. 2d. Using a mean-field formula we obtain $T_c = 58 \pm 7$ K, consistent with previous results within experimental uncertainty^{2–4}. Analysing the energy gap at different points (Fig. 2e and f), we observe a round, un-nested Fermi surface with a nearly uniform gap, making it unlikely that the gap is caused by other instabilities such as charge density waves.

We now turn to the most unexpected aspect of our 1UC data: the clearly resolved replica bands labelled A' and B' in Fig. 1d–f. Aside from an approximately 100-meV relative shift in energy, all features of A are

replicated in A'. This includes the backbending near k_F , seen in Fig. 3. Similarly B and B' are also separated by the same energy shift with nearly the same dispersion. In Fig. 1d we mark the peaks associated with A, A' and B, B' blue and red, respectively, to illustrate their presence even in the raw EDC data. In the Supplementary Information we rule out the possibility that such replicas could be caused by band structure effects (such as quantum well states^{17,18}) and further discuss weaker features such as C and D'.

We believe the replica bands are due to the shaking off of quanta of the bosonic modes in SrTiO_3 , reminiscent of the vibron shake-offs in the photoemission spectra of H_2 molecules¹⁹. In the Supplementary Information we make the case that such bosonic modes are associated with a high-energy SrTiO_3 oxygen phonon band^{5,6} near 100 meV, although its exact energy may be modified somewhat by the presence of the overlaid film. This identification is supported by recent high-resolution angle-resolved photoemission spectroscopy (ARPES) on SrTiO_3 surface states, which shows a phonon-induced hump at approximately 100 meV away from the main band²⁰ and through inelastic neutron scattering⁷. Resolving replicas of an entire band with such a clear dispersion, as seen in Fig. 1d, is unprecedented in a solid. This is made possible here by the substantial electron–phonon coupling, the fact that the collective mode energy is greater than the width of the electron band below E_F , and most importantly, the fact that the electron–phonon coupling allows only a small momentum transfer to the electron (see below and Supplementary Information).

We now turn to the multi-unit-cell films. We have measured 2UC and 30UC FeSe films grown on similar Nb-doped SrTiO_3 substrates. In Fig. 3a–f we compare the second energy derivative of the ARPES spectra for the 1UC (Fig. 3a and d), 2UC (Fig. 3b and e) and 30UC (Fig. 3c and f) films. We find that the 2UC and 30UC band structure consists of both electron-like and hole-like bands crossing E_F near M, similar to what has been observed for bulk FeSe²¹. This is very different from the 1UC film's band structure, where only electron bands cross E_F . This implies that the 1UC film is much more heavily electron-doped than even the 2UC film⁴. Most importantly, we observe neither a superconducting-like energy gap nor replica bands in the multi-UC films. A complete comparison of the band structures can be found in Extended Data Fig. 2. The fact that we find only the 1UC film to have the superconducting gap is consistent with the conclusion of ref. 1, where despite transport data showing multi-UC films to be superconducting, it is argued that only the 1UC portion of the film actually superconducts.

The temperature evolution of the replica bands in the 1UC film is plotted in Fig. 3g–j (additional temperatures are plotted in Extended Data Fig. 3). We see that these replica bands, and hence this anomalous electron–phonon coupling, persist to temperatures significantly above the gap-opening temperature. The contrast between the 1UC and multi-UC

¹Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. ²Departments of Physics and Applied Physics, and Geballe Laboratory for Advanced Materials, Stanford University, Stanford, California 94305, USA. ³Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. ⁴Quantum Matter Institute, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. ⁵Department of Physics and Astronomy, University of Tennessee, Knoxville, Tennessee 37996-1200, USA. ⁶Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. ⁷Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁸Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA. ⁹Material Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

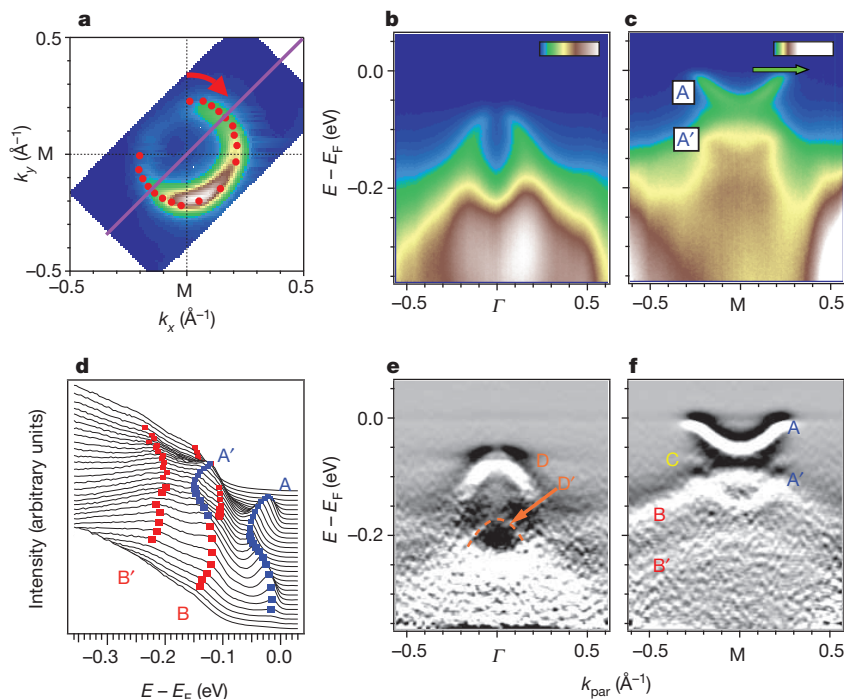


Figure 1 | Fermi surface map and high symmetry cuts of 1UC FeSe on SrTiO₃. **a**, Plot of the Fermi surface with only electron pockets located at the zone corner (M-point; k_x and k_y refer to momentum directions in the Brillouin zone). All momentum axes in this paper are given in units of inverse angstroms relative to the high-symmetry points. Red dots are the approximate points on the Fermi surface where gaps were extracted in Fig. 2e, f. **b**, **c**, High-symmetry cuts along the purple line plotted in **a**, taken at 16 K. **b** and **c** are centred at Γ (zone centre) and M, respectively (k_{par} refers to momentum along the Γ -M direction). The hole band seen in **b** is located 80 meV below E_F . In **c** a different colour scale highlights two important features: the electron band with a minimum at 60 meV below E_F (labelled A), and a

replica electron band (labelled A'), which is located 100 meV below the former and sits on top of a broad hole band. **d**, EDCs at M shown as a waterfall plot, with markers indicating band peaks. **e**, **f**, Second derivatives in energy of the high-symmetry cuts from **b** and **c**. An additional weaker replica, labelled C, can now be seen at M in **f**, sitting below A, and at the Γ point in **e** we see the hole band and a corresponding replica, labelled D and D', respectively. Data for all figures were obtained using samples grown with the same method, as described in the Supplementary Information. Spectra shown in all figures are representative data taken from six different samples that yielded comparable results.

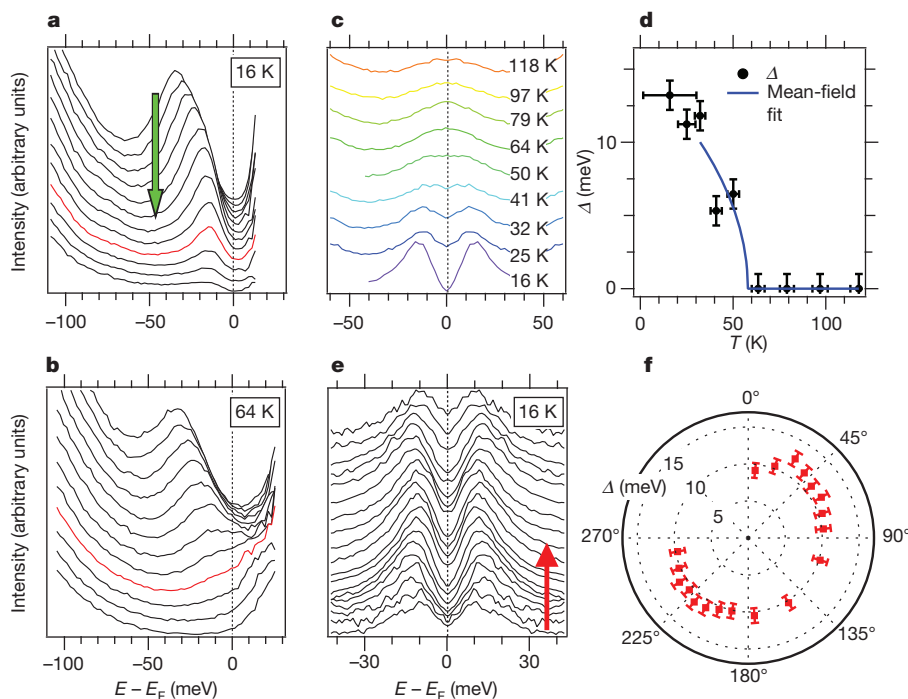
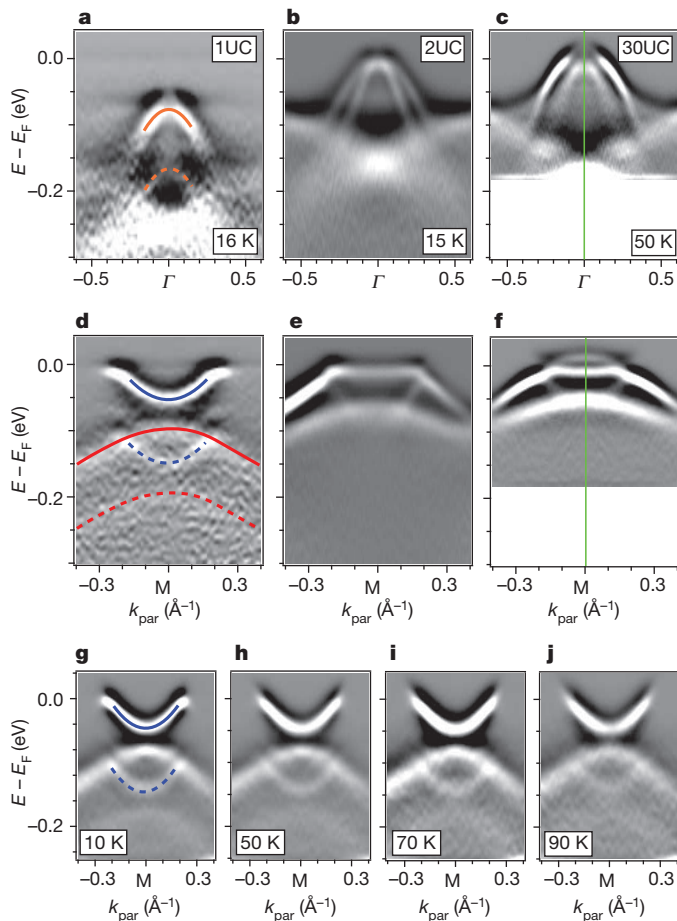


Figure 2 | Temperature dependence of the 1UC film energy gap at the M-point. **a**, **b**, Plots of the Γ -M high-symmetry dispersion EDCs at 16 K and 64 K respectively, with the Fermi-Dirac distribution factored out. The cut direction is given by the green arrow marked in Fig. 1c. The red trace represents the EDC at the Fermi momentum k_F . **c**, Plot of the evolution of the symmetrized EDCs at k_F as a function of temperature, where we observe a gap closing between 50 K and 64 K. **d**, The gap evolution as a function of temperature, with gap magnitude Δ extracted using the model in ref. 16. Error bars in energy include the drift of E_F as measured relative to a gold reference. Error bars in temperature include temperature differences within the sample manipulator. A fit to a mean-field type of order parameter is plotted, giving a gap closing temperature of 58 ± 7 K, with the range representing the uncertainty in fitting parameters. **e**, Symmetrized EDCs at different k_F values along the Fermi surface at 16 K, in the direction indicated by the red arrow in Fig. 1a. **f**, Polar plot of the gap of the EDCs from **e**. Error bars again include the drift of E_F as measured relative to a gold reference.



films suggests that the SrTiO_3 phonon which causes the replica bands is also responsible for enhancing Cooper pairing²². The fact that the replica bands in the 1UC film follows the dispersion of the main band so

Figure 3 | Dependence of electronic structure on FeSe film thickness.

a, Spectrum of the 1UC film at Γ . **b**, Γ spectrum for the 2UC film. **c**, Γ spectrum for the 30UC film. **d**, Spectrum of the 1UC film at M. **e**, M spectrum for the 2UC film. **f**, M spectrum for the 30UC film. Data in **c** and **f** have been symmetrized around Γ and M, respectively, as indicated by the green line, while data on the 1UC and 2UC films are unsymmetrized. The coloured lines in **a** and **d** are guides to the eye, with solid lines denoting the main band and dashed lines corresponding to the replicas. The main bands and replicas are colour-coded according to Fig. 1. We do not observe replica bands for either the 2UC and 30UC films. Raw spectra can be found in Extended Data Fig. 2.

g–j, Temperature dependence of the replica bands, which persist at temperatures greater than the gap-opening temperature.

closely suggests that, upon either absorption or emission, the phonons can transmit only small momenta to the electron. Such a strong forward-focused electron–phonon interaction can enhance Cooper pairing in most symmetry channels, including those with a sign change (see Supplementary Information and Extended Data Fig. 6)^{8–14}.

To estimate the strength of the electron–phonon coupling, we perform a high-statistics scan at M at low temperature, plotted in Fig. 4a. Using a spline background, we find a lower bound of 1/6 for the intensity ratio of the replica band to the main band (see Fig. 4b and Extended Data Fig. 4). We then take the intensity ratio as input and theoretically estimate the electron–phonon coupling strength (see Supplementary Information). Plotted in Fig. 4c is a simulated spectral function calculated using a model in which both the electron and hole bands couple to a flat phonon band with energy 80 meV. By tuning the coupling strength and the forward-focusing parameter we can reproduce a band–replica separation of approximately 100 meV. The simulated EDC is plotted in Fig. 4b. The ARPES spectrum is well reproduced, especially the abrupt loss in spectral weight of band A' beyond a certain momentum window, and the momentum broadening of the bands, plotted in Extended Data Fig. 5.

From the intensity ratios we obtain the electron–phonon coupling constant $\lambda \approx 0.5$, which is substantial considering that only a narrow range of phonon modes at such high frequency contribute to the coupling (owing to nearly forward scattering). This estimate in turn yields an effective phonon-mediated attraction strength $v_{\text{eff}} \approx 10$ meV (see Supplementary Information). Under the assumption that Cooper pairing is

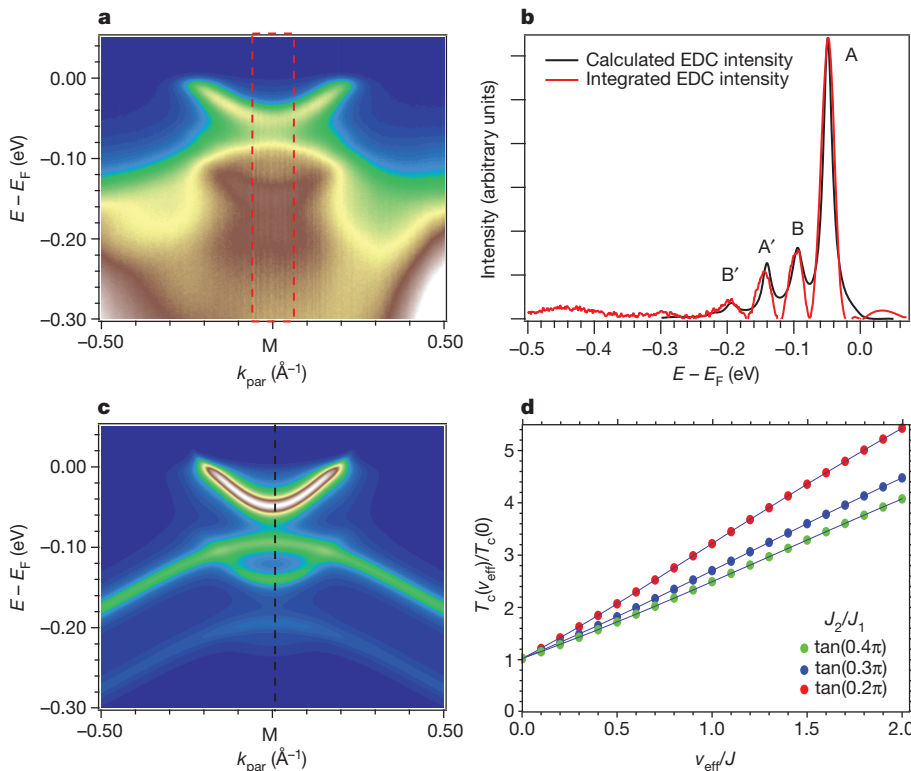


Figure 4 | Extraction of the electron-phonon coupling and determination of T_c enhancement.

a, High-statistics scan at M taken at low temperature (10 K). The spectral weight is integrated over the momentum range indicated by the dotted rectangle to obtain better statistics for a single EDC. **b**, The integrated EDC at M (after background subtraction) and the EDC from our calculations. Peaks corresponding to the bands are labelled according to Fig. 1a. **c**, Model spectral function calculation including both hole and electron bands coupled to a dispersionless 80 meV phonon mode (see Supplementary Information for details). The black dotted line indicates the EDC plotted in **b**. **d**, Plot of the T_c enhancement as a function of effective attractive electron–electron interaction strength (v_{eff}/J). Shown are plots for three different values of J_2/J_1 . The parameters we used to construct this curve (see Supplementary Information for additional details) are $q_0 = 0.1\pi/a$ ($a = 3.9$ Å), $J = 30$ meV, cutoff energy = 65 meV, and $T_c = 40$ K in the absence of electron–phonon interaction. Using the extracted parameters, we place the lower bound of the enhancement factor at 1.5.

caused by the magnetic interaction in the absence of the electron–phonon interaction, and following the effective Hamiltonian approach described in ref. 23, we first determine the pairing symmetry without the electron–phonon interaction as a function of the ratio between the nearest- and second-neighbour magnetic exchange constants J_1 and J_2 . The band structure we use is plotted in Extended Data Fig. 7 (ref. 22). We note that for the physically relevant J_2/J_1 ratio (>0.5) the pairing is an in-phase s -wave between the two electron pockets^{24,25}; see Extended Data Fig. 8. We then determine the enhancement of the Cooper pairing temperature as a function of the ratio between v_{eff} and the antiferromagnetic exchange constant $J = \sqrt{(J_1^2 + J_2^2)}$ (see Supplementary Information). We obtain $J \approx 30$ meV from the largest nearest- and second-neighbour exchange constants determined from neutron scattering experiments known to us²⁶, thus overestimating the magnetic coupling. Figure 4d plots the T_c enhancement as a function of v_{eff}/J for varying ratios of J_2 to J_1 . Using the v_{eff} extracted from our data and the above estimate of J , we determine the enhancement factor to be about 1.5. This enhancement is a lower bound, because we use the most conservative estimate of the electron–phonon coupling, and the largest J value. For bulk materials with similar band structures, for example, $\text{K}_x\text{Fe}_{2-y}\text{Se}_2$, (ref. 27), or materials from which we obtain the above stated J value²⁶, the T_c values range from 30 K to 40 K. Multiplying by the enhancement factor yields a gap-opening temperature in fairly good agreement with our films. It should be noted that we do not use the T_c of bulk FeSe because our films are heavily electron doped and hence have a very different band structure from that of bulk FeSe.

Motivated by the physical picture presented here, we propose a heterostructure where 1UC FeSe is sandwiched between SrTiO_3 on both sides, effectively doubling v_{eff} . A simple reading of Fig. 4d suggests a T_c enhancement of around 2.5, placing the Cooper pairing temperature well above the temperature of liquid nitrogen.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May; accepted 25 September 2014.

- Wang, Q. Y. *et al.* Interface induced high temperature superconductivity in single unit-cell FeSe films on SrTiO_3 . *Chin. Phys. Lett.* **29**, 037402 (2012).
- Liu, D. F. *et al.* Electronic origin of high-temperature superconductivity in single-layer FeSe superconductor. *Nature Commun.* **3**, 931 (2012).
- He, S. *et al.* Phase diagram and electronic indication of high-temperature superconductivity at 65K in single-layer FeSe films. *Nature Mater.* **12**, 605–610 (2013).
- Tan, S. *et al.* Interface-induced superconductivity and strain-dependent spin density waves in FeSe/ SrTiO_3 thin films. *Nature Mater.* **12**, 634–640 (2013).
- Choudhury, N., Walter, E. J., Kolesnikov, A. I. & Loong, C.-K. Large phonon band gap in SrTiO_3 and the vibrational signatures of ferroelectricity in ATiO_3 perovskites: first-principles lattice dynamics and inelastic neutron scattering. *Phys. Rev. B* **77**, 134111 (2008).
- Lasota, C., Wang, C.-Z., Yu, R. & Krakauer, H. Ab initio linear response study of SrTiO_3 . *Ferroelectrics* **194**, 109–118 (1997).
- Stirling, W. G. Neutron inelastic scattering study of the lattice dynamics of strontium titanate: harmonic models. *J. Phys. C* **5**, 2711–2730 (1972).
- Bulut, N. & Scalapino, D. J. $d_{x^2-y^2}$ symmetry and the pairing mechanism. *Phys. Rev. B* **54**, 14971–14973 (1996).
- Johnston, S. *et al.* Specific heat of $\text{Ca}_{0.32}\text{Na}_{0.68}\text{Fe}_2\text{As}_2$ single crystals: unconventional \pm multiband superconductivity with intermediate repulsive

interband coupling and sizable attractive intraband couplings. *Phys. Rev. B* **89**, 134507 (2014).

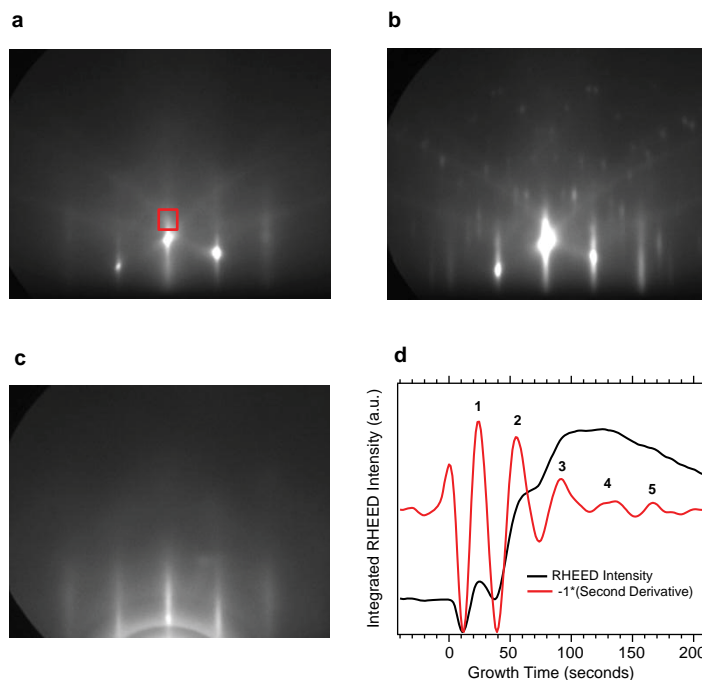
- Sandvik, A. W., Scalapino, D. J. & Bickers, N. E. Effect of an electron-phonon interaction on the one-electron spectral weight of a d -wave superconductor. *Phys. Rev. B* **69**, 094523 (2004).
- Johnston, S. *et al.* Systematic study of electron-phonon coupling to oxygen modes across the cuprates. *Phys. Rev. B* **82**, 064513 (2010).
- Varelogiannis, G., Perali, A., Cappelluti, E. & Pietronero, L. Density-of-states-driven anisotropies induced by momentum decoupling in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$. *Phys. Rev. B* **54**, R6877–R6880 (1996).
- Aperis, A., Kotetes, P., Varelogiannis, G. & Oppeneer, P. M. Small- q phonon-mediated unconventional superconductivity in the iron pnictides. *Phys. Rev. B* **83**, 092505 (2011).
- Santi, G., Jarlborg, T., Peter, M. & Weger, M. s - and d -wave symmetries of the solutions of the Eliashberg equations. *Physica C* **259**, 253–264 (1996).
- Hsu, F. C. *et al.* Superconductivity in the PbO-type structure α -FeSe. *Proc. Natl Acad. Sci. USA* **105**, 14262–14264 (2008).
- Norman, M. R., Randeria, M., Ding, H. & Campuzano, J. C. Phenomenology of the low-energy spectral function in high- T_c superconductors. *Phys. Rev. B* **57**, R11093–R11096 (1998).
- Pagel, J. J., Miller, T. & Chiang, T. C. Occupied and unoccupied band structure of Ag(100) determined by photoemission from Ag quantum wells and bulk samples. *Phys. Rev. B* **61**, 1804–1810 (2000).
- Tang, S. J., Lee, Y. R., Chang, S. L., Miller, T. & Chiang, T. C. Umklapp-mediated quantization of electronic state in Ag films on Ge(111). *Phys. Rev. Lett.* **96**, 216803 (2006).
- Turner, D. *Molecular Photoelectron Spectroscopy* (Wiley, 1970).
- Chang, Y. J., Bostwick, A., Kim, Y. S., Horn, K. & Rotenberg, E. Structure and correlation effects in semiconducting SrTiO_3 . *Phys. Rev. B* **81**, 235109 (2010).
- Maletz, J. *et al.* Unusual band renormalization in the simplest iron based superconductor. Preprint at <http://arxiv.org/abs/1307.1280> (2013).
- Xiang, Y.-Y., Wang, F., Wang, D., Wang, Q.-H. & Lee, D.-H. High-temperature superconductivity at the FeSe/ SrTiO_3 interface. *Phys. Rev. B* **86**, 134508 (2012).
- Davis, J. C. S. & Lee, D.-H. Concepts relating magnetic interactions, intertwined electronic orders, and strongly correlated superconductivity. *Proc. Natl Acad. Sci. USA* **110**, 17623–17630 (2013).
- Yang, F., Wang, F. & Lee, D.-H. Fermiology, orbital order, orbital fluctuations, and Cooper pairing in iron-based superconductors. *Phys. Rev. B* **88**, 100504 (2013).
- Yu, R., Goswami, P., Si, Q., Nikolic, P. & Zhu, J.-X. Superconductivity at the border of electron localization and itinerancy. *Nature Commun.* **4**, 2783 (2013).
- Wang, M. *et al.* Doping dependence of spin excitations and its correlations with high-temperature superconductivity in iron pnictides. *Nature Commun.* **4**, 2874 (2013).
- Zhang, Y. *et al.* Nodeless superconducting gap in $\text{A}_x\text{Fe}_2\text{Se}_2$ ($\text{A} = \text{K}, \text{Cs}$) revealed by angle-resolved photoemission spectroscopy. *Nature Mater.* **10**, 273–277 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division. D.-H.L. is supported by the Department of Energy, Office of Basic Energy Sciences, Division of Materials Science, under the Quantum Material programme DE-AC02-05CH11231. Measurements were performed at the Stanford Synchrotron Radiation Lightsource, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences.

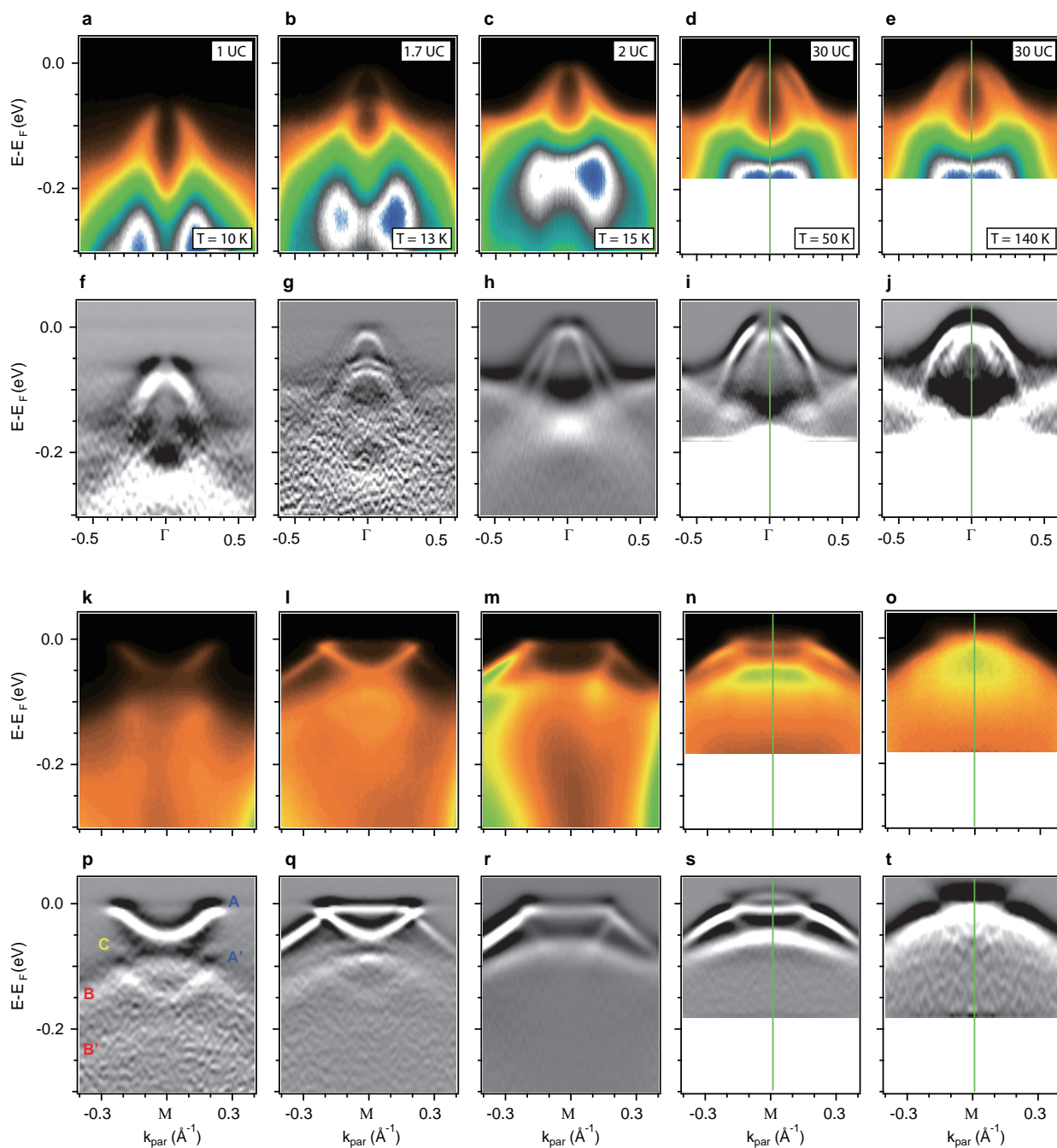
Author Contributions J.J.L., F.T.S. and R.G.M. grew films, collected and analysed data, and wrote the paper. S.J. and D.-H.L. performed theory calculations. Y.T.C., W.L., Z.K.L., Y.Z., D.H.L. and M.Y. provided discussion about data and interpretation. M.H. and D.H.L. provided experimental support at Stanford Synchrotron Radiation Lightsource. All authors participated in the discussion of results. Project direction was provided by D.-H.L., T.P.D. and Z.-X.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.-X.S. (zxshen@stanford.edu).



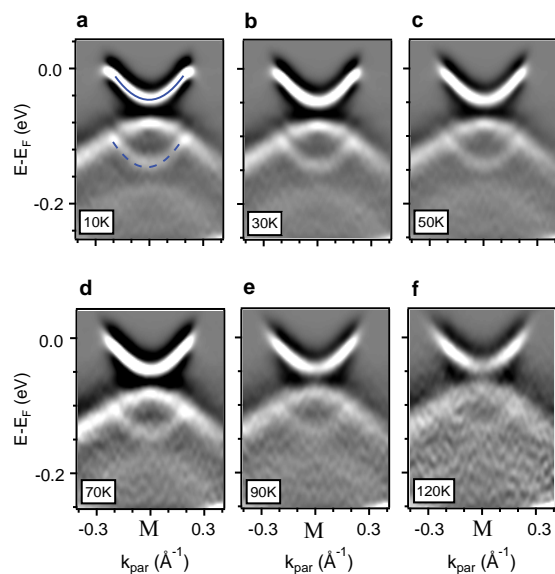
Extended Data Figure 1 | Reflection high-energy electron diffraction (RHEED) images observed during FeSe growth. **a**, RHEED image of SrTiO_3 substrate after degassing at 450°C for 1 h. Red box highlights the region integrated for monitoring RHEED oscillations. **b**, Surface reconstruction as observed by RHEED at annealing temperatures. **c**, RHEED image of FeSe 1 UC

film showing uniform streaks typical of an atomically flat thin film. **d**, RHEED intensity for integration region shown in **a** (black). The second derivative of the intensity curve (red), with numbers indicating the number of layers grown highlights the RHEED oscillations signalling the completion of a unit cell after about 30 s.

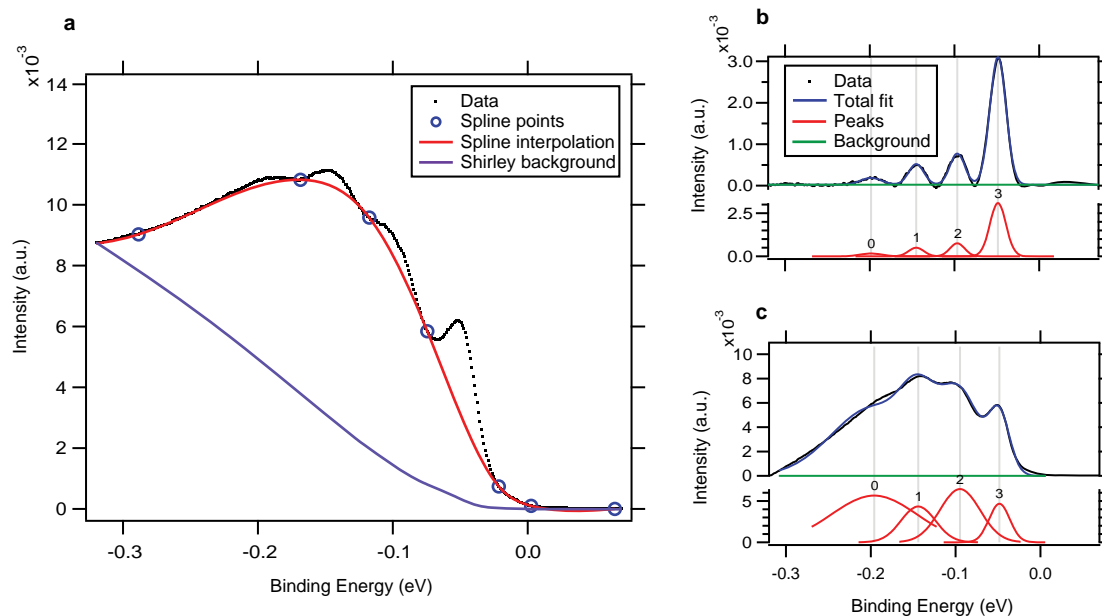


Extended Data Figure 2 | Raw spectra and second derivatives of 1UC, 1.7UC, 2UC and 30UC films. **a, f, k, p,** Plots of the raw spectrum at Γ , the second derivative at Γ , the raw spectrum at M, and the second derivative at M for the 1UC film, respectively, taken at 10 K. **b, g, l, q,** The same plots for the 1.7UC film taken at 13 K. **c, h, m, r,** The same plots for the 2UC film taken at 15 K. **d, i, n, s,** The same plots for the 30UC film taken at 50 K. **e, j, o, t,** The same

plots for the 30UC film taken at 140 K. Data for the 30UC film are symmetrized about the high-symmetry points (indicated by the green line), and were taken with 25-eV photons. We observe a band splitting in the 30UC film at M, at low temperature (s). This band splitting closes at higher temperature (140 K), where we now observe only one band (t).

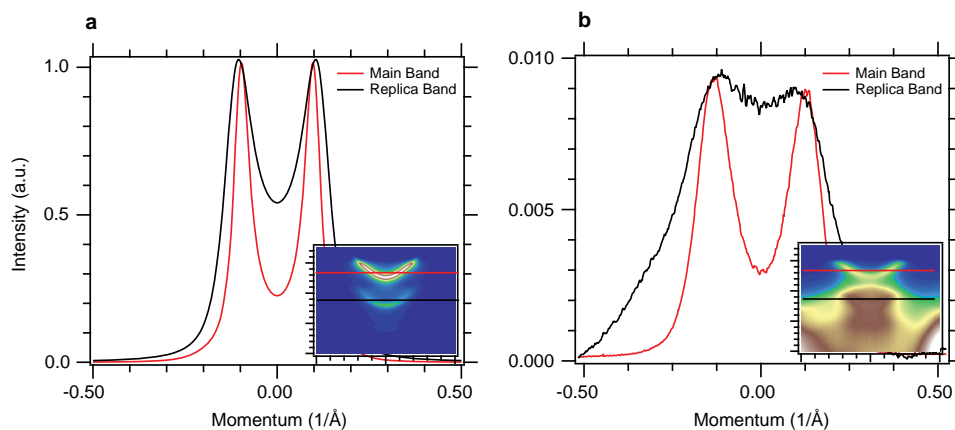


Extended Data Figure 3 | Temperature evolution of the M point spectrum of the 1UC film. **a**, Spectrum at 10 K, where we can clearly see the backbending of the replica bands, exactly like the main bands near E_F . **b**, Spectrum at 30 K. **c**, Spectrum at 50 K. **d**, Spectrum at 70 K. **e**, Spectrum at 90 K. **f**, Spectrum at 120 K. The replica bands persist up to temperatures significantly higher than the gap-opening temperature.



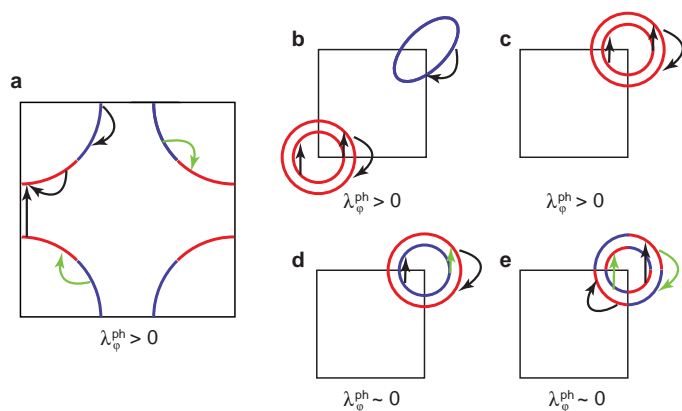
Extended Data Figure 4 | Fitting the intensities of the ARPES spectra at M.
a, Plot showing two different backgrounds used in the fitting. Using the blue circles as fixed points, we first modelled the background using a spline interpolation, plotted in red. The second fit used a Shirley background (see Supplementary Information), plotted in purple. **b,** Data and fitting with the

spline background subtracted. We fitted to four Gaussian peaks, which are plotted separately for clarity. **c,** Data and fitting with the Shirley background subtracted. We restrict our fitting energy window to be from -0.32 eV to 0.03 eV.

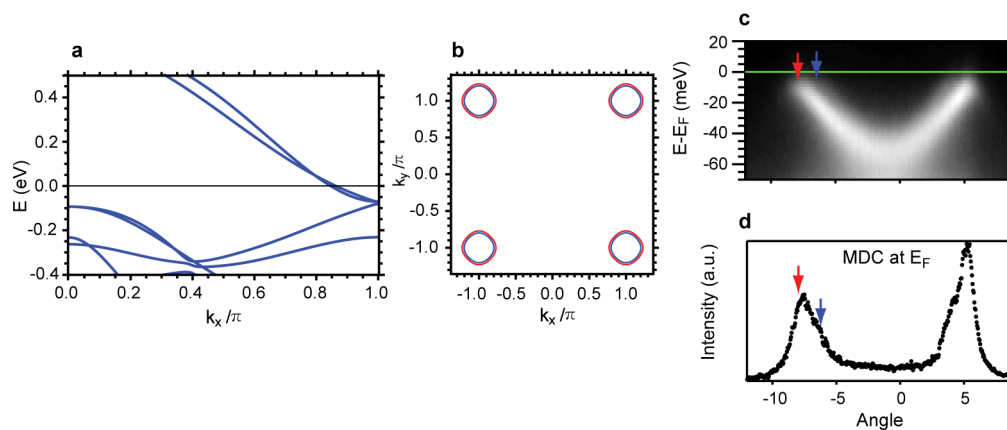


Extended Data Figure 5 | Momentum distribution comparison between the main band and replica band. **a**, The momentum distribution curves (MDCs) of our theoretical calculation of the main electron band and the replica electron band with normalized intensities. The MDCs of both bands are taken at the same energies with respect to their band bottoms (see inset). The replica

band peaks are broadened due to the electron–phonon coupling. **b**, The MDCs of our data, with a momentum-independent background subtracted from the replica band MDC. The momentum-dependent background—such as contributions from the hole band—is the likely cause of the extra broadening in the data.

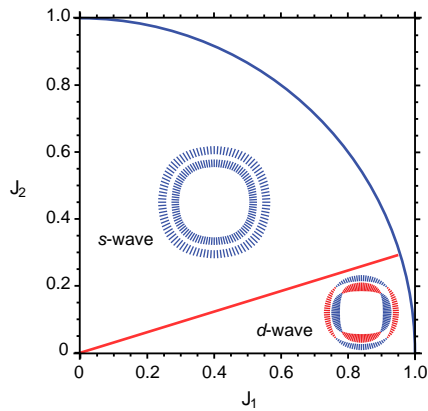


Extended Data Figure 6 | Effects of electron-phonon coupling on different gap symmetries and Fermi surfaces. Cartoon sketches of the various Fermi surfaces and gap symmetries found or proposed in unconventional superconductors. $\lambda_{\phi}^{\text{ph}}$ is the projected coupling defined by equation (18) in the Supplementary Information. **a**, Sketch for the copper oxides with a *d*-wave gap. **b–e**, Various scenarios for the iron-based superconductors. Only one-quarter of the first Brillouin zone is shown for clarity. The thick blue and red lines indicate the phase of the gap. The arrows show various forward-focused scattering processes. The black arrows indicate scattering processes that connect portions of the Fermi surface with the same sign gap and are therefore pair-enhancing. The green arrows show pair-breaking processes which connect regions of the Fermi surface with different signs.



Extended Data Figure 7 | Input electronic structure for calculated T_c enhancement. **a**, Calculated band structure used in our determination of T_c enhancement. **b**, Calculated Fermi surface showing slightly split electrons

pockets. **c**, Dispersion along the M-point showing two nearly degenerate bands indicated with red and blue arrows. **d**, Momentum distribution curve (MDC) at E_F showing peaks from the two bands plotted in **c**.



Extended Data Figure 8 | Phase diagram of the $J_1 - J_2$ model. The blue line represents $\sqrt{J_1^2 + J_2^2} = 1$. The red line represents the transition between different gap symmetries at $J_2/J_1 \approx 0.31$. Above the transition one finds a gap with *s*-wave symmetry. Below the transition one finds a gap with *d*-wave symmetry. Diagrams of the two different possible symmetries are drawn in their respective region of the phase diagram as insets. The lengths of the tick lines in the inset diagrams represent the magnitude of the gap, while the colour represents the sign (red for minus, blue for plus). The two electron pockets in the figure are separated for clarity.

Formic-acid-induced depolymerization of oxidized lignin to aromatics

Alireza Rahimi¹, Arne Ulbrich¹, Joshua J. Coon^{1,2} & Shannon S. Stahl¹

Lignin is a heterogeneous aromatic biopolymer that accounts for nearly 30% of the organic carbon on Earth¹ and is one of the few renewable sources of aromatic chemicals². As the most recalcitrant of the three components of lignocellulosic biomass (cellulose, hemicellulose and lignin)³, lignin has been treated as a waste product in the pulp and paper industry, where it is burned to supply energy and recover pulping chemicals in the operation of paper mills⁴. Extraction of higher value from lignin is increasingly recognized as being crucial to the economic viability of integrated biorefineries^{5,6}. Depolymerization is an important starting point for many lignin valorization strategies, because it could generate valuable aromatic chemicals and/or provide a source of low-molecular-mass feedstocks suitable for downstream processing⁷. Commercial precedents show that certain types of lignin (lignosulphonates) may be converted into vanillin and other marketable products^{8,9}, but new technologies are needed to enhance the lignin value chain. The complex, irregular structure of lignin complicates chemical conversion efforts, and known depolymerization methods typically afford ill-defined products in low yields (that is, less than 10–20wt%)^{2,10,11}. Here we describe a method for the depolymerization of oxidized lignin under mild conditions in aqueous formic acid that results in more than 60wt% yield of low-molecular-mass aromatics. We present the discovery of this facile C–O cleavage method, its application to aspen lignin depolymerization, and mechanistic insights into the reaction. The broader implications of these results for lignin conversion and biomass refining are also considered.

Lignin is biosynthesized from a relatively small number of phenylpropanoid building blocks that contribute to recurring structural elements in this otherwise complex polymeric material. One of the most common motifs is the ‘β-O-4’ alkyl–aryl ether linkage between aromatic rings², which features a secondary benzylic alcohol at the Cα position and a primary aliphatic alcohol at the Cγ position (Fig. 1). Fundamental

studies of wood pulp delignification in the context of paper production have shown that oxidation of the Cα alcohol to a ketone (for example by using stoichiometric Mn or Cr oxide reagents) facilitates the removal of lignin from cellulose, apparently by promoting cleavage of the β-O-4 linkage^{12,13}. This concept has not been applied to the production of low-molecular-mass aromatics from lignin. Several groups, including our own, have probed the reactions of oxidized lignin model compounds, such as **2** (Fig. 1b), with varying degrees of success^{12,14–21}. The recent development of a method for the chemoselective aerobic oxidation of Cα alcohols to ketones in native lignin, which achieved ~90% conversion to ‘lignin^{ox}’, enables direct investigation of the reactivity of oxidized lignin. (Oxidative transformations of lignin and lignin model compounds have been the focus of extensive study and are summarized in refs 2, 20, 21. Primary references associated with this work are compiled in Supplementary Information.)

We observed previously that the oxidized lignin model **2** reacts with alkaline hydrogen peroxide to yield the aromatic monomers veratric acid (88% yield) and guaiacol (42%)¹⁴. The instability of the guaiacol under the reaction conditions prompted us to consider reductive cleavage methods. The reactivity of **2** was tested in the presence of different reducing metals, including zinc, aluminium, magnesium, iron and manganese, in aqueous formic acid at 110 °C (Fig. 2a). Four different major products were identified from these reaction conditions. In the presence of zinc, **2** afforded small amounts of the O-formylated product **3** (6%), together with good yields of the aryl ethyl ketone reductive cleavage product **4** (76%) and guaiacol (69%) (Fig. 2a, entry 1). When the reaction was performed with metal sources other than Zn (Fig. 2a, entries 2–5), the diketone product **5** was observed instead of **4**. Good yields of **5** (74%) and guaiacol (63%) were obtained from the reaction with manganese. Numerous other conditions were tested with these metal sources (see Supplementary Table 1); however, no further improvement in the product yields was observed. For example, substantially lower

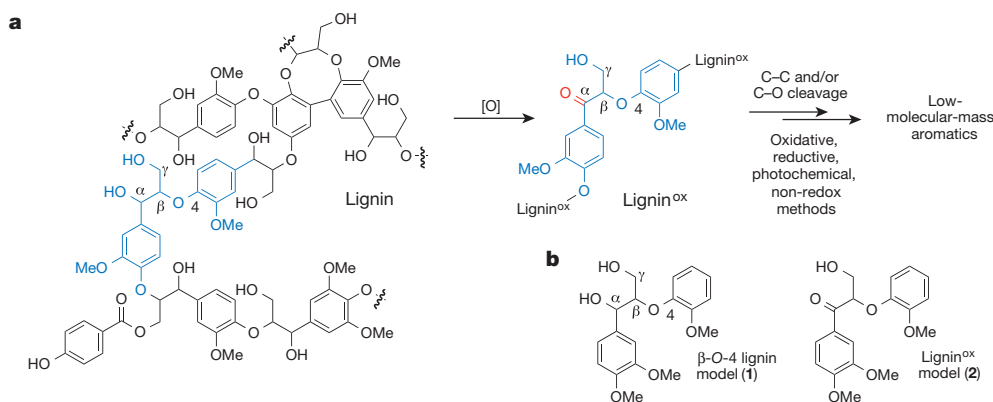


Figure 1 | Lignin structure and strategies for depolymerization.

a, Representative structure of a fragment of poplar (including aspen) lignin, highlighting a β-O-4 unit, together with the strategy for lignin conversion to

low-molecular-mass aromatics by chemoselective alcohol oxidation, followed by C–C and/or C–O cleavage. **b**, Structure of β-O-4 model compounds **1** and **2**.

¹Department of Chemistry, University of Wisconsin–Madison, 1101 University Avenue, Madison, Wisconsin 53706, USA. ²Department of Biomolecular Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA.

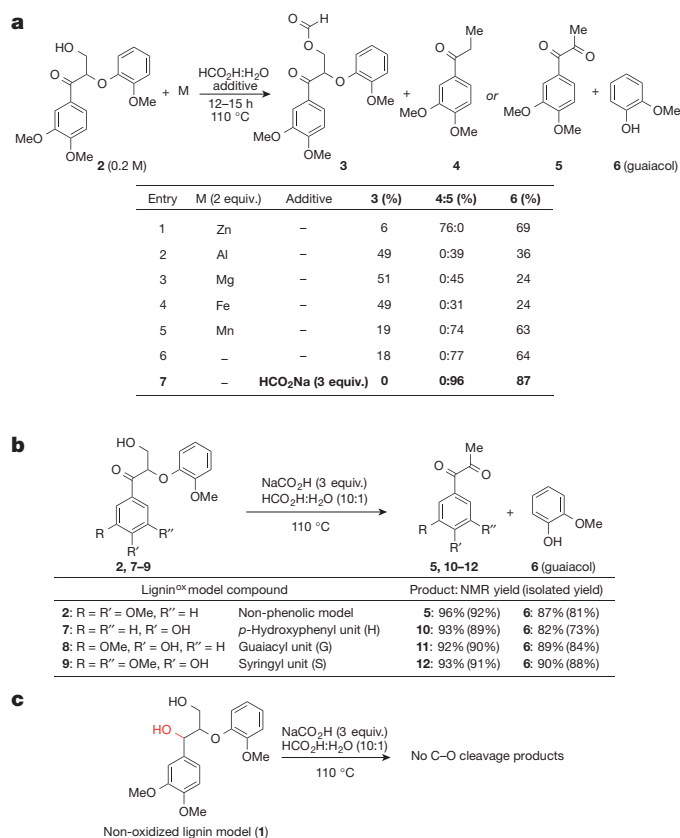


Figure 2 | C–O cleavage of lignin model compounds with formic acid.

a, Cleavage products obtained from the treatment of lignin^{ox} model compound **2** in formic acid with and without reducing metals (M). Yields shown in the table were determined by ¹H NMR spectroscopy, using hexamethyldisiloxane as the internal standard. **b**, Cleavage of different lignin^{ox} model compounds in the presence of formic acid and sodium formate. **c**, Control reaction with the non-oxidized lignin model compound **1** under the formic acid/formate cleavage conditions.

yields were obtained with organic solvents (ethanol, γ -valerolactone or toluene) or on replacement of formic acid with another acid source (acetic acid, HCl or H₂SO₄).

The conversion of **2** into the diketone product **5** is a redox-neutral process. We therefore tested the reaction in the absence of a reducing metal, and products **5** and **6** were obtained in yields comparable to the best results obtained with a stoichiometric metal source (Fig. 2a, entry 6). Even better yields of **5** and **6** resulted (96% and 87%, respectively) when 3 equivalents of sodium formate were included in the reaction mixture (Fig. 2a, entry 7). These conditions proved to be effective with several other lignin^{ox} model compounds, including those derived from *p*-hydroxyphenyl (H)-, guaiacyl (G)- and syringyl (S)-type lignin units 7–9 (Fig. 2b). The C α ketone is crucial to the success of the reaction: no cleavage products were observed when the non-oxidized model compound **1** was subjected to the formic acid/formate reaction conditions (Fig. 2c).

These results provided the basis for testing the reactivity of authentic lignin polymer. Aspen is a representative hardwood, and a sample of native aspen lignin was isolated and determined to have a ratio of S:G subunits of 2.2:1 on the basis of two-dimensional heteronuclear single quantum coherence NMR analysis (see Supplementary Fig. 1)²². After treatment of this material under the aerobic conditions described previously¹⁴, the oxidized lignin was subjected to the formic acid/sodium formate reaction conditions at 110 °C. The amount of sodium formate added to the reaction mixture was estimated from the mass of the lignin sample to provide about 3 equivalents of formate per S/G aromatic subunit. The formic acid was evaporated after 24 h, and the residue was extracted

with ethyl acetate. A soluble fraction, corresponding to 61.2wt% of the original lignin, was obtained, together with an insoluble fraction that accounted for 29.7wt% of the lignin (Fig. 3a). Acetylation of the insoluble material and analysis by gel-permeation chromatography revealed a significant decrease in molecular mass relative to the initial lignin polymer (Fig. 3b). The soluble fraction, containing the majority of the lignin-derived mass, was analysed by high-resolution mass spectrometry coupled with liquid chromatography (LC–MS) to identify and quantify the products of the reaction (Fig. 3c and Supplementary Table 2). Product identities were confirmed by comparison with authentic samples obtained commercially or synthesized independently (see Supplementary Information for details).

Overall, more than 52% of the original lignin was converted to well-defined aromatic compounds (Fig. 3c). Syringyl and guaiacyl-derived diketones, directly analogous to those observed in the model study (compare Fig. 2b), are two of the major depolymerization products (19.8%). The S:G diketone ratio (2:1) is very similar to the S:G monomer composition in the lignin. Additional syringyl and guaiacyl-derived aromatics (for example, syringaldehyde and vanillin) account for most of the remaining product mass, together with *p*-hydroxybenzoic acid (4%), which is probably derived from *p*-hydroxybenzoic esters present in aspen (and other poplar) lignins²³. A full listing of characterized products is provided in Supplementary Table 2.

An attempt to use unoxidized lignin in the reaction resulted in only 7.2wt% yield of low-molecular-mass aromatics. The products that are observed probably arise from the small amount of C α ketones present in native lignin (evident from two-dimensional NMR spectroscopy). This

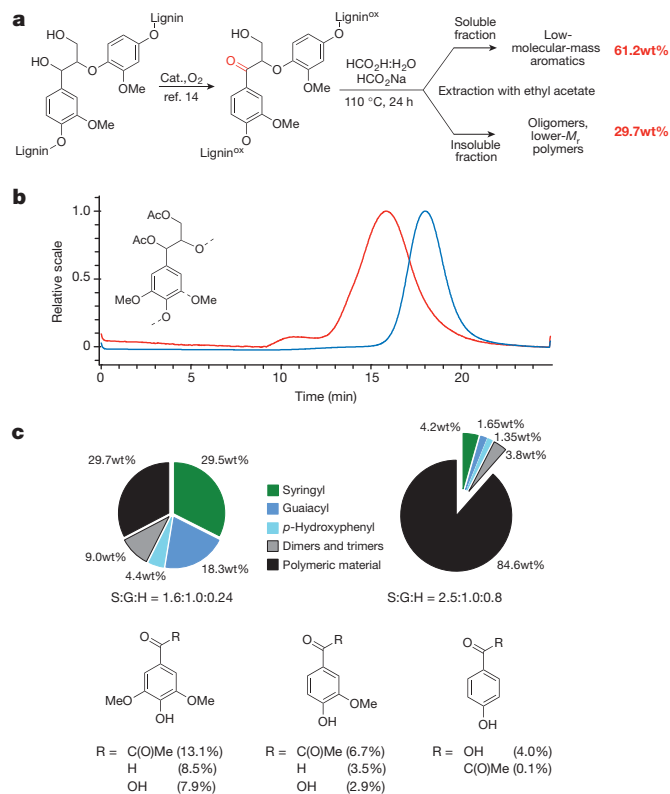


Figure 3 | Depolymerization of aspen lignin with formic acid.

a, Depolymerization of oxidized aspen lignin. **b**, Gel-permeation chromatography of the insoluble fraction from depolymerization (29.7wt%) (blue trace) compared with that of native lignin (red trace). For the insoluble fraction of depolymerized lignin, $M_w/M_n = 1.08$ and $M_n = 4,600$ Da; for native lignin, $M_w/M_n = 1.18$ and $M_n = 10,800$ Da. **c**, Identification, quantification and distribution of products obtained from the depolymerization of oxidized lignin (upper left) compared with those from native lignin (upper right). Bottom: major depolymerization products from oxidized lignin. For a full product listing see Supplementary Table 2.

result demonstrates the importance of C α oxidation in promoting lignin depolymerization and provides strong motivation to develop improved lignin oxidation methods suitable for large-scale application. The C α benzylic alcohol in lignin is activated electronically relative to the C γ primary alcohol, and our previous study showed that numerous classes of oxidants promote the chemoselective oxidation of the C α alcohol¹⁴. These observations bode well for the development of improved oxidation methods.

Mechanistic insights into the depolymerization process were obtained from additional studies of model compound **2**. Reaction time-course data, obtained by ¹H NMR spectroscopy, revealed that **2** converts quickly into the alcohol formylation product **3**, followed by a slower conversion of **3** into products **5** and **6** (Fig. 4a, b). Conversion of **2** into **3** is rapid, even at room temperature (Fig. 4c). Direct investigation of the reaction of **3** revealed a small steady-state concentration of the aryl vinyl ether intermediate **13** during the formation of diketone product **5** (Fig. 4d). Independent testing of **13** confirmed that this intermediate is consumed more rapidly than the overall conversion of **3** into **5** (Fig. 4e). A large deuterium kinetic isotope effect, observed when the independent reaction rates of **2** and **2**-d₁ were compared ($k_{\text{H}}/k_{\text{D}} = 9.3 \pm 0.2$; Fig. 4f), indicates that the elimination of formic acid from **3** is the rate-limiting step in the transformation. This kinetic isotope effect is larger than the semi-classical limit associated with the cleavage of a C–H bond, but it aligns

with a previous study providing evidence for proton tunnelling in concerted E2 elimination reactions²⁴.

Overall, the sequence is a redox-neutral process that results in no net consumption of formic acid (Fig. 4g). This feature distinguishes the present approach from other lignin conversion methods that employ formic acid as a source of H₂ in transfer hydrogenation/hydrogenolysis reactions with heterogeneous catalysts (see, for example, refs 25, 26). According to the mechanism in Fig. 4g, the beneficial effect of lignin oxidation may be attributed to the ability of the benzylic carbonyl group to polarize the C–H bond and lower the barrier for the rate-limiting E2 elimination reaction. Both the increased C–H acidity and orbital overlap between the carbonyl C=O π system and the developing π bond of the alkene are expected to contribute to a lower barrier for this step. The mechanism also accounts for the beneficial effect of using a ‘buffered’ reaction medium, containing both formate and formic acid (compare Fig. 2a, entries 6 and 7). The rate-limiting elimination step in Fig. 4g is proposed to involve both a base (formate) to remove the proton and an acid (formic acid) to assist in the loss of the formate as a leaving group.

The overall yield of structurally identified, monomeric aromatics obtained here is the highest reported so far for lignin depolymerization^{2,10,11}. Although some of the products obtained from this process have direct commercial value (for example vanillin and syringaldehyde)^{27,28}, the more important result may be the generation of a stream of soluble aromatic

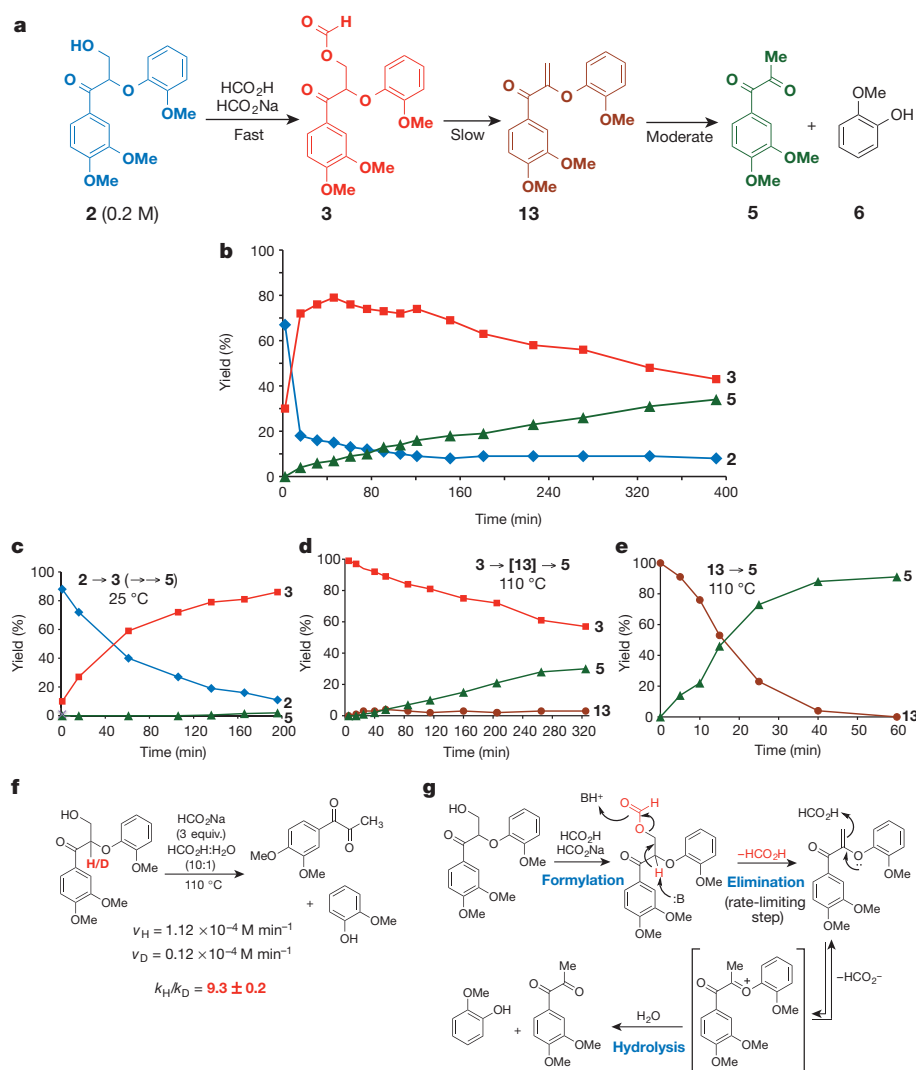


Figure 4 | Mechanistic study of C–O cleavage in **2.** **a, b**, Different intermediates (**a**) and overall time course of the reaction **2** → **3** → [**13**] → **5** at 110 °C (**b**). **c–e**, Time courses of formylation of **2** at 25 °C (**c**), elimination of

formic acid from **3** at 110 °C (**d**), and hydrolysis of aryl vinyl ether **13** at 110 °C (**e**). **f**, Kinetic isotope effect for cleavage of the oxidized model compound **2**. **g**, Proposed mechanism for C–O cleavage.

feedstocks for further upgrading. For example, low-molecular-mass feedstocks should reduce coking and char formation, enhance conversions and facilitate product separations in processes with heterogeneous catalysts (such as hydrogenation, hydrogenolysis, decarbonylation and decarboxylation). The syringyl, guaiacyl and *p*-hydroxyphenyl aromatic products arise directly from the monomeric composition of the lignin, and their yields correlate closely with the quantity of β -O-4 linkages present in the original lignin. These observations highlight the importance of developing chemical conversion technologies for S-, G- and H-derived aromatics and suggest that plants containing lignin with high β -O-4 content (as much as 85% has been observed²⁹) could be particularly appealing feedstocks for biomass valorization. This work further draws attention to biomass separation methods. Since the inception of the pulp and paper industry, biomass separation methods have emphasized the production of high-purity cellulose. Most existing process conditions, ranging from the classic kraft pulping to modern organosolvolytic methods, significantly modify or damage the lignin and lead to a significant decrease or complete loss of β -O-4 structural units. Biomass separation methods that deliver both high-purity sugar and native-type lignin streams, such as a recent γ -valerolactone-based process³⁰, could gain a competitive advantage in biorefining applications.

Received 24 June; accepted 15 September 2014.

Published online 2 November 2014.

- Ralph, J., Brunow, G. & Boerjan, W. Lignins. *eLS* <http://dx.doi.org/10.1002/9780470015902.a0020104> (2007).
- Zakzeski, J., Bruijninx, P. C. A., Jongerius, A. L. & Weckhuysen, B. M. The catalytic valorization of lignin for the production of renewable chemicals. *Chem. Rev.* **110**, 3552–3599 (2010).
- Sanderson, K. Lignocellulose: a chewy problem. *Nature* **474**, S12–S14 (2011).
- Lin, S. Y. & Lin, I. S. in *Ullmann's Encyclopedia of Industrial Chemistry* 5th edn (eds Elvers, B., Hawkins, S., Schulz, G.) Vol. A15, 305–315 (VCH, 1985).
- Tuck, C. O., Pérez, E., Horváth, I. T., Sheldon, R. A. & Poliakoff, M. Valorization of biomass: deriving more value from waste. *Science* **337**, 695–699 (2012).
- Holladay, J. E., Bozell, J. J., White, J. F. & Johnson, D. *Top Value-added Chemicals from Biomass* Vol. 2 <http://www1.eere.energy.gov/bioenergy/pdfs/pnnl-16983.pdf> (United States Department of Energy, 2007).
- Ragauskas, A. J. *et al.* Lignin valorization: improving lignin processing in the biorefinery. *Science* **344**, 1246843 (2014).
- Hocking, M. B. Vanillin: synthetic flavoring from spent sulfite liquor. *J. Chem. Educ.* **74**, 1055–1059 (1997).
- Business Areas, *Borregaard* <http://www.borregaard.com/Business-Areas> (20 June 2014).
- Pandey, M. P. & Kim, C. S. Lignin depolymerization and conversion: a review of thermochemical methods. *Chem. Eng. Technol.* **34**, 29–41 (2011).
- Wang, H., Tucker, M. & Ji, Y. Recent development in chemical depolymerization of lignin: a review. *J. Appl. Chem.* **2013**, 1–9 (2013).
- Gierer, J. & Nore'n, I. Oxidative pretreatment of pine wood to facilitate delignification during kraft pulping. *Holzforschung* **36**, 123–130 (1982).
- Ljunggren, S. & Olsson, A. The specificity in oxidation of some lignin and carbohydrate models and pine wood shavings with permanganate and pyridinium dichloromate before the kraft pulping process. *Holzforschung* **38**, 91–99 (1984).
- Rahimi, A., Azarpira, A., Kim, H., Ralph, J. & Stahl, S. S. Chemoselective metal-free aerobic oxidation in lignin. *J. Am. Chem. Soc.* **135**, 6415–6418 (2013).
- Crestini, C. & D'Auria, M. Singlet oxygen in the photodegradation of lignin models. *Tetrahedron* **53**, 7877–7888 (1997).
- Fukagawa, N. & Ishizu, A. Photoreaction of phenacyl aryl ether type lignols. *J. Wood Chem. Technol.* **11**, 263–289 (1991).
- Argyropoulos, D. S. & Sun, Y. Photochemically induced solid-state degradation, condensation, and rearrangement reactions in lignin model compounds and milled wood lignin. *Photochem. Photobiol.* **64**, 510–517 (1996).
- Vanucci, C., De Violet, P. F., Bouas-Laurent, H. & Castellan, A. Photodegradation of lignin: a photophysical and photochemical study of a non-phenolic α -carbonyl β -O-4 lignin model dimer, 3,4-dimethoxy- α -(2'-methoxyphenoxy) acetophenone. *J. Photochem. Photobiol. Chem.* **41**, 251–265 (1988).
- Nguyen, J. D., Matsuura, B. S. & Stephenson, C. R. J. A photochemical strategy for lignin degradation at room temperature. *J. Am. Chem. Soc.* **136**, 1218–1221 (2014).
- Collinson, S. R. & Thielemans, W. The catalytic oxidation of biomass to new materials focusing on starch, cellulose and lignin. *Coord. Chem. Rev.* **254**, 1854–1870 (2010).
- Chatel, G. & Rogers, R. D. Oxidation of lignin using ionic liquids—an innovative strategy to produce renewable chemicals. *ACS Sustainable Chem. Eng.* **2**, 322–339 (2014).
- Chang, H., Cowling, E. B. & Brown, W. Comparative studies on cellulolytic enzyme lignin and milled wood lignin of sweetgum and spruce. *Holzforschung* **29**, 153–159 (1975).
- Ralph, J. & Landucci, L. L. in *Lignin and Lignans; Advances in Chemistry* (eds Heitner, C., Dimmel, D. R. & Schmidt, J. A.) 137–234 (CRC Press, 2010).
- Fouad, F. M. & Farrell, P. G. Primary and secondary kinetic isotope effects in E2 elimination reactions. *Tetrahedr. Lett.* **19**, 4735–4738 (1978).
- Xu, W., Miller, S. J., Agrawal, P. K. & Jones, C. W. Depolymerization and hydrodeoxygenation of switchgrass lignin with formic acid. *ChemSusChem* **5**, 667–675 (2012).
- Toledano, A. *et al.* Fractionation of organosolv lignin from olive tree clippings and its valorization to simple phenolic compounds. *ChemSusChem* **6**, 529–536 (2013).
- Bauer, K., Garber, D. & Surburg, H. in *Ullmann's Encyclopedia of Industrial Chemistry* 5th edn (eds Elvers, B., Rounsaville, J. F. & Schulz, G.) Vol. A11, 199–200 (VCH, 1985).
- Björsvik, H.-R. & Liguori, L. Organic processes to pharmaceutical chemicals based on fine chemicals from lignosulfonates. *Org. Process Res. Dev.* **6**, 279–290 (2002).
- Martínez, Á. T. *et al.* Monolignol acylation and lignin structure in some nonwoody plants: a 2D NMR study. *Phytochemistry* **69**, 2831–2843 (2008).
- Luterbacher, J. S. *et al.* Nonenzymatic sugar production from biomass using biomass-derived γ -valerolactone. *Science* **343**, 277–280 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Ralph for numerous helpful discussions, H. Kim and A. Azarpira for assistance with the purification and NMR characterization of aspen lignin, and S. Chakraborty for assistance with gel-permeation chromatographic analysis of lignin samples. Financial support for this project was provided by the Great Lakes Bioenergy Research Center (Department of Energy Biological and Environmental Research Office of Science DE-FC02-07ER64494). The NMR facility was partly supported by the National Science Foundation (CHE-9208463) and the National Institutes of Health (S10 RR08389).

Author Contributions A.R. and S.S.S. conceived the idea for the lignin depolymerisation. A.R. performed the chemical reactions of lignin and lignin model compounds, including NMR spectroscopic characterization of the products. A.U. and J.J.C. designed the analytical approach to characterize the lignin depolymerization products. A.U. performed LC/MS analysis. A.R. and S.S.S. wrote the manuscript in consultation with A.U. and J.J.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S.S. (stahl@chem.wisc.edu).

Subduction-driven recycling of continental margin lithosphere

A. Levander¹, M. J. Bezada^{2,3}, F. Niu^{1,4}, E. D. Humphreys², I. Palomeras¹, S. M. Thurner¹, J. Masy¹, M. Schmitz⁵, J. Gallart⁶, R. Carbonell⁶ & M. S. Miller⁷

Whereas subduction recycling of oceanic lithosphere is one of the central themes of plate tectonics, the recycling of continental lithosphere appears to be far more complicated and less well understood¹. Delamination and convective downwelling are two widely recognized processes invoked to explain the removal of lithospheric mantle under or adjacent to orogenic belts^{2–5}. Here we relate oceanic plate subduction to removal of adjacent continental lithosphere in certain plate tectonic settings. We have developed teleseismic body wave images from dense broadband seismic experiments that show higher than expected volumes of anomalously fast mantle associated with the subducted Atlantic slab under northeastern South America and the Alboran slab beneath the Gibraltar arc region^{6,7}; the anomalies are under, and are aligned with, the continental margins at depths greater than 200 kilometres. Rayleigh wave analysis^{8,9} finds that the lithospheric mantle under the continental margins is significantly thinner than expected, and that thin lithosphere extends from the orogens adjacent to the subduction zones inland to the edges of nearby cratonic cores. Taking these data together, here we describe a process that can lead to the loss of continental lithosphere adjacent to a subduction zone. Subducting oceanic plates can viscously entrain and remove the bottom of the continental thermal boundary layer lithosphere from adjacent continental margins. This drives surface tectonics and pre-conditions the margins for further deformation by creating topography along the lithosphere–asthenosphere boundary. This can lead to development of secondary downwellings under the continental interior, probably under both South America and the Gibraltar arc^{8,10}, and to delamination of the entire lithospheric mantle, as around the Gibraltar arc¹¹. This process reconciles numerous, sometimes mutually exclusive, geodynamic models proposed to explain the complex oceanic–continental tectonics of these subduction zones^{12–17}.

In both the southern Caribbean (CAR) and the Gibraltar arc the subduction zones propagate towards the Atlantic along Mesozoic passive margins, with the oceanic plates tearing away from the adjacent continents as they descend into the mantle. We start with the simpler subduction zone in the southeastern CAR (Fig. 1), where the southern edge of the Atlantic slab, ATL, is interacting with the northern edge of the South American continental margin. Since the Late Palaeocene, the Antilles subduction zone has traversed northern South America as the Americas have moved westward^{15,18} (Fig. 1), leaving a steeply dipping ATL slab near the southeastern corner of the Antilles subduction zone⁶. This subduction zone terminates at the El Pilar–San Sebastian strike-slip fault system in northern Venezuela, characterized as a STEP fault¹⁹. A zone of intermediate depth seismicity (less than ~150 km), the Paria cluster, is hypothesized to be the mechanical expression of the descending ATL slab tearing from the South America continental margin²⁰. Here we are concerned with what is occurring on the underside of the lithosphere where it is nearly aseismic and viscous, rather than in the brittle regime near the plate surface.

Body wave tomography shows that the fast ATL anomaly is larger and lies twice as far south of the plate boundary under northern South America⁶ than predicted by North American–South American plate motion models for the Cenozoic^{21,22}. The anomaly, well resolved in an east–west direction, extends into the mantle transition zone as a steeply dipping, fast slab (Fig. 2). Lithosphere thickness determined from surface wave tomography and Ps and Sp receiver functions in northern South America⁸ (Figs 2, 3) indicate that the continental lithosphere is ~50–100% thicker adjacent to and east of the subduction zone than to the west: east of the descending ATL plate at about 64° W, the South America lithosphere is almost uniformly ~110–120 km thick from the Guayana shield to the plate boundary (Fig. 3). To the west, along the coast that the subduction zone has traversed, South American lithosphere is 55–90 km thick between the plate boundary and the Guayana shield (64–68° W, Fig. 3). Restoring the fast ATL anomaly to the surface, the

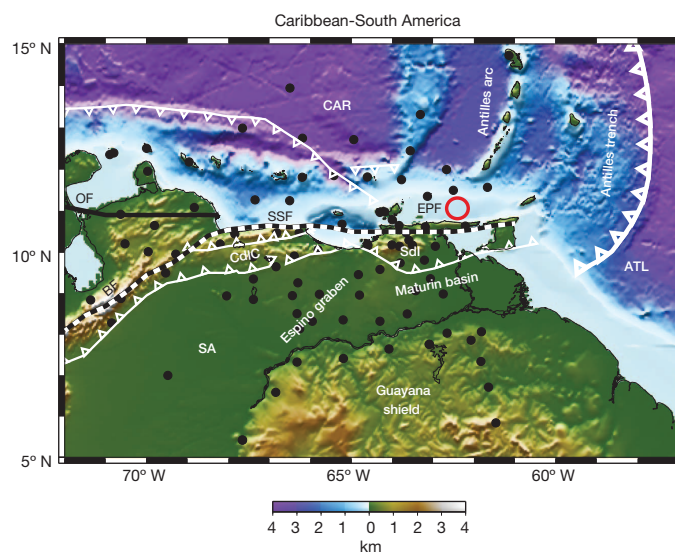


Figure 1 | Colour topography and bathymetry of the southeastern Caribbean showing plate boundaries and significant tectonic features.

The Antilles subduction zone has traversed northern South America from west to east, creating the coastal mountain belts. The Antilles subduction zone and the limits of orogenic belts are shown in heavy and light white lines with teeth indicating the upper plate. Plate and block-bounding right-lateral strike-slip faults are shown as heavy black and white dashed lines. Solid black line is an inactive strike-slip fault. Colour key shows topography and bathymetry relative to sea level. Red circle indicates the Paria seismic cluster. Black dots are locations of broadband seismograph stations. CAR, Caribbean plate; SA, South American plate; ATL, Atlantic plate attached to South America; SSF, San Sebastian fault; EPF, El Pilar fault; BF, Bocono fault; OF, Oca fault; SdI, Serrania del Interior; CdIC, Cordillera de la Costa.

¹Earth Science Department, Rice University, Houston, Texas 77005-1892, USA. ²Department of Geological Sciences, University of Oregon, Eugene, Oregon 97043, USA. ³Department of Earth Sciences, University of Minnesota, Minneapolis, Minnesota 55455-0231, USA. ⁴State Key Laboratory of Petroleum Resources and Prospecting, and Unconventional Natural Gas Institute, China, University of Petroleum, Beijing 102249, China. ⁵Fundación Venezolana de Investigaciones Sismológicas, Caracas 1073, Venezuela. ⁶Institut de Ciències de la Terra Jaume Almera, CSIC, Barcelona 08028, Spain. ⁷Department of Earth Sciences, University of Southern California, Los Angeles, California 90089-0740, USA.

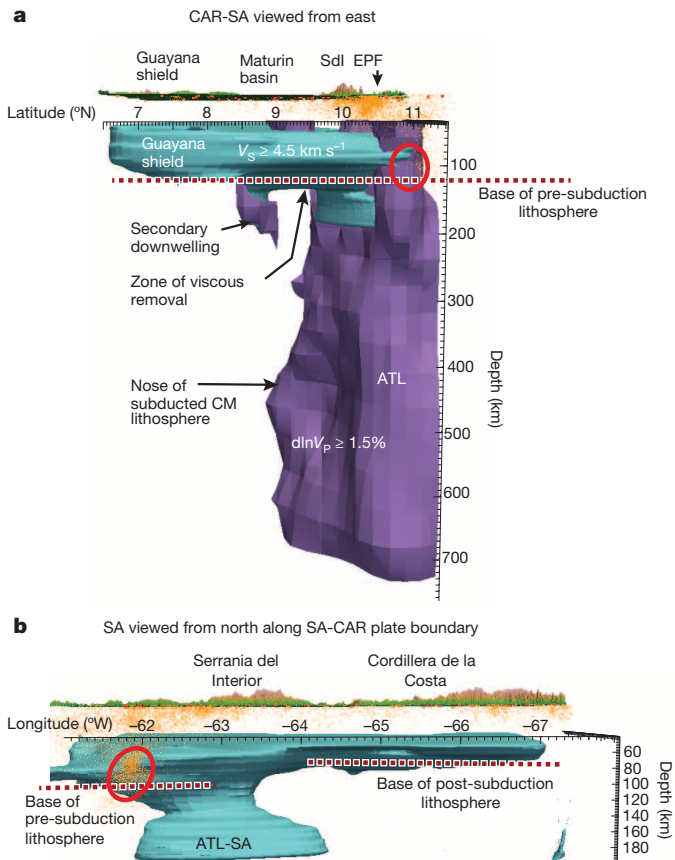


Figure 2 | Composite three-dimensional seismic images of the Atlantic plate and South American lithosphere. **a**, The composite seismic image viewed from the east showing the CAR-SA (Caribbean–South American plate boundary) and SA lithosphere. An iso-surface from a teleseismic P-wave tomography model encloses $d\ln V_p$ anomalies faster than 1.5% (purple). Another iso-velocity surface from a Rayleigh wave tomography model encloses $V_s \geq 4.50 \text{ km s}^{-1}$ (blue), which we take as a proxy for the SA lithospheric mantle. The P-wave tomography images the subducted Atlantic (ATL) slab and subducted continental mantle (CM) lithosphere. The anomaly's southern end is south of the plate-bounding El Pilar fault (EPF) and extends to the west and south as a nose of high velocity. The lithosphere thickness is $\sim 110\text{--}120 \text{ km}$ in the region between the Guayana Shield and the ATL slab (dashed red line). A secondary downwelling is imaged by both seismic methods under the Maturin Basin. Seismicity is shown as small orange dots, the red circle indicates the Paria seismic cluster. **b**, Same surface wave volume as **a**, viewing the SA-CAR boundary from the north along the Paria seismic cluster. Lithosphere thickness decreases from $\sim 100 \text{ km}$ east of the subducting slab to $\sim 75 \text{ km}$ west of this slab, as indicated by the red dotted lines. Subduction is ongoing against the SA margin east of $\sim 63^\circ$ and has progressed beyond the SA margin to the west.

slab extends several hundred kilometres underneath northern South America, suggesting that the southern edge of the slab anomaly includes South American mantle lithosphere removed from at least as far west as $\sim 69^\circ \text{ W}$ (Fig. 3). We hypothesize that the excess mass south of the plate boundary is the base of the continental margin lithosphere that was viscously removed with the ATL plate as it subducted.

The Espino graben (Fig. 1) is part of the rifted continental margin that formed along northern and eastern South America at the opening of the central Atlantic $\sim 160 \text{ Myr}$ ago. Sufficient time has elapsed since the Atlantic opening for thermal boundary layer (TBL) mantle to form to a depth of at least 100 km under the entire passive margin^{23,24}, comparable to the lithosphere thickness we observe in northeastern South America. The Espino graben lithosphere is now only $55\text{--}70 \text{ km}$ thick (Figs 2 and 3), $30\text{--}45 \text{ km}$ less than expected from TBL growth, and thinner than the lithosphere elsewhere to the west. Because rifting left the graben lithosphere

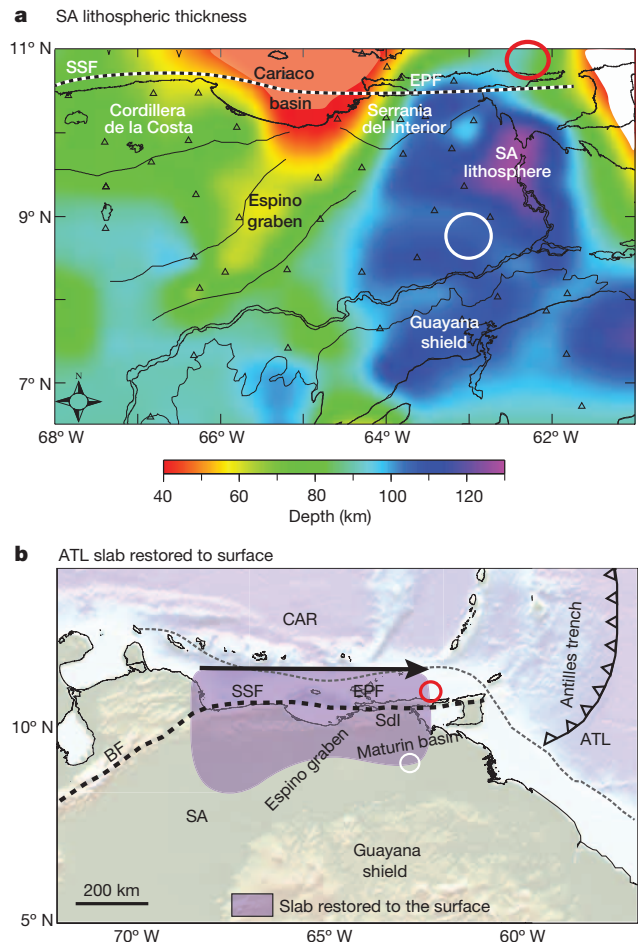


Figure 3 | Map of lithospheric thickness in northeastern South America and restoration of the Atlantic slab P-wave anomaly to the surface. **a**, The lithosphere thickness is less than 100 km everywhere west of about $\sim 64^\circ$, and north of the Guayana shield, and thicker than 100 km to the east. Red circle indicates the Paria seismic cluster, white circle the location of the secondary downwelling. Abbreviations as Fig. 1, black dashed lines are active plate and block bounding strike-slip faults. Triangles are broadband seismograph stations. **b**, The ATL slab anomaly restored to the surface. The restored anomaly underlies twice as much of the South American continent as North America–South America plate convergence predicts. The thick dashed line is the South America–Caribbean plate boundary, the thin dashed line is location of the modern passive margin off eastern South America and the nominal location of the palaeo-passive margin off northern South America. Arrow indicates the direction of subduction zone motion relative to South America. Arrowhead terminates at the Paria cluster, where the ATL is currently tearing from South America.

thinner than along other parts of the margin, it grew a greater thickness of TBL lithosphere than elsewhere. This lithosphere has now been removed partly or entirely. To us this suggests that the continental lithospheric mantle removed by the descending ATL plate is TBL lithosphere, rather than chemically depleted mantle.

An additional deep high-velocity anomaly $\sim 250 \text{ km}$ south of the subducting ATL (near 8.75° N) has been imaged by body and surface wave tomography (Fig. 2). It extends to $\sim 200 \text{ km}$ depth beneath the southern Maturin basin, the foreland basin of the coastal fold belt. We suggest that this anomaly is a convective instability that developed in response to mantle flow closer to the subduction zone. Instabilities in the continental interior can result from hydration and weakening of the continental lithosphere from exposure to mantle wedge asthenosphere flowing from the subduction zone, combined with continental margin loading by the foreland basin, orogenic belt and descending ATL. Flexure due to loading can drive lower crustal mafic lithologies into eclogite facies,

creating a high-density anomaly in the lower crust. We suggest that viscous removal of TBL mantle during subduction has helped to trigger convective instability under the continental interior.

The Gibraltar arc is part of a diffuse plate boundary that extends north across the Atlas Mountains, Alboran Sea and Betic Mountains (Fig. 4). Slab rollback, slab breakoff, lithospheric downwelling and lithospheric delamination have been invoked to explain different aspects of the Gibraltar arc structure^{12,14,17,25–29}. Prior to the initiation of western Mediterranean subduction tectonics in the Oligocene, both southern Iberia and northwestern Africa were passive margins. About 18 Myr of subduction rollback of east-dipping Alpine Tethys seafloor between the two continents (Fig. 4) has left an oceanic slab suspended in the upper mantle beneath the Gibraltar arc^{7,29} (Extended Data Fig. 1), the remnant of Neogene slab rollback^{25,26}. Although the details are still debated, mantle tomography supports a history in which rollback swept across the Alboran Sea region with rapid east-to-west propagation along the African margin while slowing along the Iberian margin, so that the subduction zone rotated clockwise as it propagated westward towards and then under the Betics, the Rif Mountains, and Gibraltar (Fig. 3). The Alboran crust was drawn westward above the retreating subduction zone⁷. The simultaneous occurrence of rollback and Eurasian–African convergence led to highly oblique thrusting of Alboran crust over the African and Iberian margins. Below the crustal overthrusting, slab rollback requires that the subducting ocean lithosphere detach from the passive margins along the edges of the Iberian and African plates.

The almost vertically hanging Alboran slab is an arcuate feature, concave southeast, underlying both the Iberian and African continental margins at 70–100 km depth, and extending through the 410-km discontinuity⁷ (Extended Data Fig. 1). Interpretation of GPS data²⁸ and a variety of Rif seismic images^{9,11} (Extended Data Fig. 2) shows that the Alboran slab is attached to and delaminating from the Rif lithosphere under northern Morocco, the African element of the Gibraltar arc. The seismic measurements detect thicker Rif crust than is predicted from local elevation.

Receiver function images show the top of the Alboran slab merging with the Rif Moho at ~50–55 km depth, evidence that the descending Alboran plate is attached to the Rif lithosphere at the Moho. Under southern Iberia the Alboran slab lies close to the base of the Betic crust on an east–west axis. Here geodetic data, receiver functions, surface wave tomography, and seismicity all suggest that the Alboran slab is still detaching or has recently detached from the base of the Betics^{9,11,30} (Extended Data Fig. 3). Thus the continental margins are in places still attached to the subducting slab on either side of the Alboran Sea.

Rayleigh wave tomography shows thin continental margin lithosphere (50–70 km) around the margins of the Alboran Sea surrounding the descending slab⁹ (Fig. 4, Extended Data Fig. 1), with thin Iberian lithosphere under the Betics parallel to the direction of slab retreat. Similar to the descending ATL slab, the Alboran slab under southern Iberia is imaged as a high-velocity curtain extending through the transition zone. Restored to the surface, this curtain occupies the Alboran Sea and the adjacent continental margins⁷ (Fig. 4). We suggest that in addition to oceanic lithosphere, the high-velocity anomaly includes continental margin lithosphere drawn from under Spain and Africa.

The structure and tectonics of North Africa are complex, as different tectonic domains interact. The thin lithosphere along the Rif margin extends into the Middle and High Atlas Mountains, where the lithosphere is only slightly thicker than the crust (Fig. 4). The High Atlas, an inverted Triassic–Jurassic graben adjacent to the Sahara craton, has anomalously thin crust for its elevation³¹. Plio–Pleistocene uplift of the High Atlas is attributed to piecemeal convective downwellings enabled by mantle flow from the Canary plume and/or around the Alboran slab³⁰ that undermined the mantle lithosphere of the Middle and High Atlas and produced intermittent basaltic volcanism. Sites of thin lithosphere in Spain also have associated basaltic volcanism, with a range of chemistries indicative of a hydrated mantle wedge, or a dry asthenospheric source. We suggest that the Alboran subduction has removed continental margin lithosphere from the base of Iberia and North Africa, creating

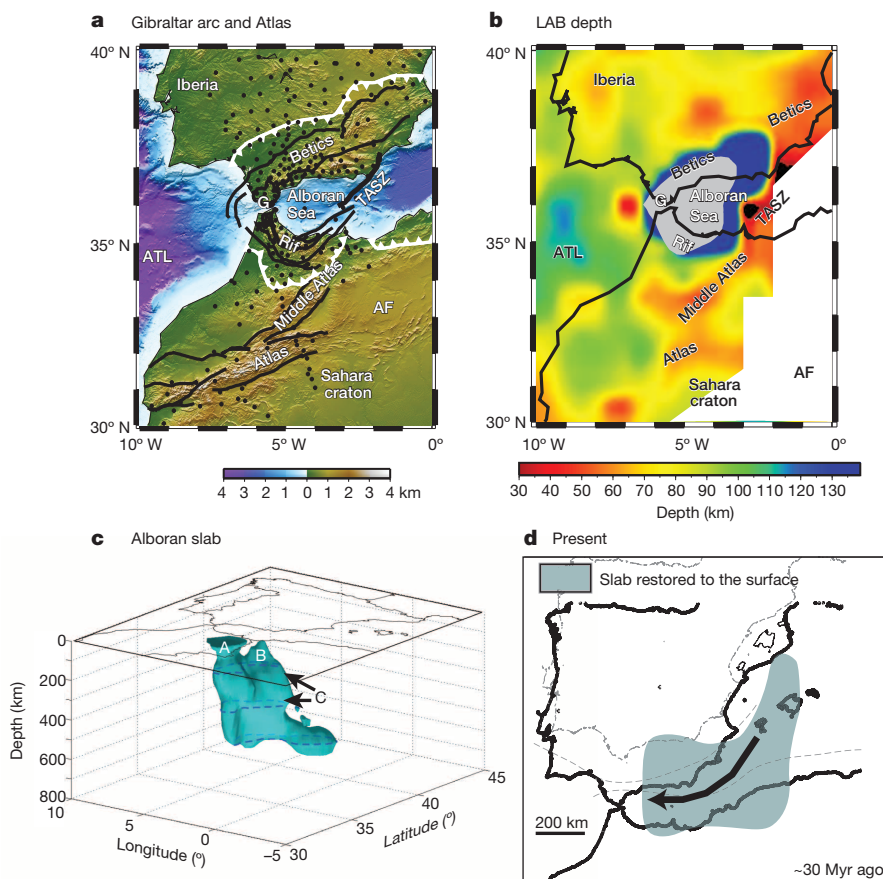


Figure 4 | Gibraltar arc region tectonic features, lithosphere thickness, positive P-wave tomography anomalies, and P-wave anomaly restored to the surface. **a**, The Gibraltar arc region is defined by the arcuate Rif–Betic mountain system. The diffuse plate boundary separating Eurasia (Iberia) from Africa (AF) extends from the Betics across Gibraltar (G), the Alboran Sea and Rif Mountains to the southern edge of the Atlas Mountains. Slab rollback proceeded from approximately northeast to west (see **d**). Limits of Betic and Rif thrust system are shown in white. ATL, Atlantic plate. TASZ, left-lateral Trans-Alboran shear zone. Black dots are broadband seismograph stations. Solid black lines are faults and geologic province boundaries **b**, Map of lithospheric thickness in the Gibraltar arc region. The lithosphere is thin north, east and south of the Alboran slab, and thinnest to the northeast and east of the slab beneath the Betics and the Alboran slab along the direction of trench retreat. **c**, The slab anomaly is shown as an iso-velocity surface enclosing $\ln V_p \geq 1.5\%$. The anomaly extends downward beneath the Rif (A) and Betics (B) and dips east from Gibraltar under the Alboran Sea. A curtain of high velocity underlies southern Spain (C). **d**, The Alboran slab anomaly restored to the surface. Dashed lines show Spain 30 Myr ago with nominal Iberia and African passive margins. Arrow indicates the direction of subduction zone motion relative to Spain's current position.

lithosphere–asthenosphere boundary (LAB) topography that helped generate secondary downwellings interior to both continental margins.

Both the southeastern Caribbean and the Gibraltar arc subduction zones are adjacent to former Mesozoic passive margins that were largely unaffected by tectonic activity between the time of rifting and the arrival of the modern subduction zones. Following rifting, like the developing ocean basin lithosphere, the passive margins form a TBL lithosphere of thickness comparable to oceanic lithosphere, that is, ~ 100 km, within ~ 70 Myr (refs 23, 24). The seismic images from northeastern South America and the Gibraltar arc demonstrate that the adjacent continental margin lithosphere is thinner than expected along the margins traversed by the subduction zones. We attribute this to convective removal of the continental margin mantle lithosphere by the adjacent subducting oceanic plate. In the Gibraltar arc this process extends to levels as shallow as the Moho, resulting in lithospheric delamination beneath both adjacent continental margins. We speculate that removal of lithospheric mantle to within a plate thickness of the continental margin creates gradients in the LAB that result in additional downwellings under the continental interior. We have imaged one such downwelling beneath the South America continent, and infer that previous downwellings in the South America interior removed lithosphere north of, but close to the Guayana shield. Slab rollback and lithosphere removal from the Gibraltar arc margins occurred shortly before large-scale downwellings beneath the Middle and High Atlas, suggesting a relation between the two. Potentially hotter asthenosphere and asthenospheric flow around the narrow Alboran slab resulted in delamination and widespread volcanism.

Loss of mantle lithosphere to downwellings and delamination beneath growing and collapsing mountain belts is well recognized locally in the Andes and in the western United States as an indirect consequence of subduction and its aftermath^{2–5}. What we describe is more similar to ‘ablative subduction’ predicted by two-dimensional geodynamic modelling³², in which a subducting oceanic plate viscously removes the TBL mantle on an overriding continental plate. Here, oceanic plate subduction, the primary convection system, viscously removes the TBL from beneath an adjacent continental margin, creating LAB topography under the continent that can give rise to secondary, edge convection³³. The mantle lithosphere that remains beneath the margins after removal of the basal TBL is likely to be the buoyant, basalt-depleted peridotite that forms the chemical boundary layer of the continental mantle lithosphere.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 September 2013; accepted 19 September 2014.

- Gao, S. *et al.* Recycling deep cratonic lithosphere and generation of intraplate magmatism in the North China Craton. *Earth Planet. Sci. Lett.* **270**, 41–53 (2008).
- Houseman, G. & Molnar, P. in *Continental Reactivation and Reworking* (eds Miller, J. A., Holdsworth, R. E., Buick, I. S. & Hand, M.) 13–38 (Spec. Publ. 184, Geological Society of London, 2001).
- Zandt, G. *et al.* Active foundering of a continental arc root beneath the southern Sierra Nevada in California. *Nature* **431**, 41–46 (2004).
- Kay, R. W. & Kay, S. M. Delamination and delamination magmatism. *Tectonophysics* **219**, 177–189 (1993).
- Levander, A. *et al.* Continuing Colorado plateau uplift by delamination-style convective lithospheric downwelling. *Nature* **472**, 461–465 (2011).
- Bezada, M. J., Levander, A. & Schmandt, B. Subduction in the southern Caribbean: Images from finite-frequency P wave tomography. *J. Geophys. Res.* **115**, B12333 <http://dx.doi.org/10.1029/2010JB007682> (2010).
- Bezada, M. J. *et al.* Evidence for slab rollback in westernmost Mediterranean from improved upper mantle imaging. *Earth Planet. Sci. Lett.* **368**, 51–60 (2013).
- Masy, J., Levander, A., Niu, F. & Schmitz, M. Lithospheric expressions of Cenozoic subduction, Mesozoic rifting and the Precambrian Shield in Venezuela. *Earth Planet. Sci. Lett.* (in the press).
- Palomeras, I., Levander, A., Liu, K., Thurner, V. A. & Gallart, J. Finite-frequency Rayleigh wave tomography of the western Mediterranean. *Geochim. Geophys. Res.* **15**, 140–160 (2014).
- Bezada, M. J., Humphreys, E. D., Palomeras, I., Levander, A. & Carbonell, R. Piecewise delamination of Moroccan lithosphere from beneath the Atlas Mountains. *Geochim. Geophys. Res.* **15**, 975–985 (2014).

- Thurner, S. M., Palomeras, I., Levander, A., Carbonell, R. & Lee, C. T. A. Ongoing lithospheric removal in the Western Mediterranean: Ps receiver function results from the PICASSO project. *Geochim. Geophys. Res.* **15**, 1113–1127 (2014).
- Loneragan, L. & White, N. Origin of the Betic-Rif mountain belt. *Tectonics* **16**, 504–522 (1997).
- Platt, J. P. & Vissers, R. L. M. Extensional collapse of thickened continental lithosphere: a working hypothesis for the Alboran Sea and Gibraltar arc. *Geology* **17**, 540–543 (1989).
- Duggen, S., Hoernle, K., Bogaard, P. d. & Harris, C. Magmatic evolution of the Alboran region: the role of subduction in forming the western Mediterranean and causing the Messinian Salinity Crisis. *Earth Planet. Sci. Lett.* **218**, 91–108 (2004).
- Pindell, J. & Kennan, L. in *The Geology and Evolution of the Region Between North and South America* (eds James, K., Lorente, M. A. & Pindell, J.) 1–55 (Spec. Publ. 328, Geological Society of London, 2009).
- VanDecar, J. C., Russo, R. M., James, D. E., Ambeg, W. B. & Franke, M. Aseismic continuation of the Lesser Antilles slab beneath continental South America. *J. Geophys. Res.* **108**, 2043 <http://dx.doi.org/10.1029/2001JB000884> (2003).
- Calvert, A., Sandvol, E., Seber, D. & Barazangi, M. Geodynamic evolution of the lithosphere and upper mantle beneath the Alboran region of the western Mediterranean: constraints from travel time tomography. *J. Geophys. Res.* **105**, 10871–10898 (2000).
- Burke, K. Tectonic evolution of the Caribbean. *Annu. Rev. Earth Planet. Sci.* **16**, 201–230 (1988).
- Govers, R. & Wortel, M. J. R. Lithosphere tearing at STEP faults: response to edges of subduction zones. *Earth Planet. Sci. Lett.* **236**, 505–523 (2005).
- Clark, S. *et al.* Eastern Venezuelan tectonics driven by lithospheric tear geodynamics, not oblique collision. *Geochim. Geophys. Res.* **9**, Q11004 <http://dx.doi.org/10.1029/2008GC002084> (2008).
- Pindell, J., Kennan, L., Stanek, K.-P., Maresch, W. & Draper, G. Foundations of Gulf of Mexico and Caribbean evolution: eight controversies resolved. *Geol. Acta* **4**, 303–341 (2006).
- Müller, R. D., Royer, J.-Y., Cande, S. C., Roest, W. R. & Maschenkov, S. in *Sedimentary Basins of the World Vol. 4* (ed. Mann, P.) 33–59 (Elsevier, 1999).
- Sleep, N. H. Thermal effects of the formation of Atlantic continental margins by continental break up. *Geophys. J. R. Astron. Soc.* **24**, 325–350 (1971).
- Sclater, J. G., Parsons, B. & Jaupart, C. Oceans and continents: similarities and differences in the mechanisms of heat loss. *J. Geophys. Res.* **86**, 11535–11552 (1981).
- Rosenbaum, G. *Tectonic Reconstruction of the Alpine Orogen in the Western Mediterranean Region* PhD thesis, Monash Univ. (2003).
- Royden, L. H. Evolution of retreating subduction boundaries formed during continental collision. *Tectonics* **12**, 629–638 (1993).
- Seber, D., Barazangi, M., Ibembrahim, A. & Demnati, A. Geophysical evidence for lithospheric delamination beneath the Alboran Sea and Rif-Betic mountains. *Nature* **379**, 785–790 (1996).
- Pérouse, E., Vernant, P., Chéry, J., Reilinger, R. & McClusky, S. Active surface deformation and sub-lithospheric processes in the western Mediterranean constrained by numerical models. *Geology* **38**, 823–826 (2010).
- Gutscher, M.-A. *et al.* Evidence for active subduction beneath Gibraltar. *Geology* **30**, 1071–1074 (2002).
- Mancilla, F. L. *et al.* Delamination in the Betic Range: deep structure, seismicity, and GPS motion. *Geology* **41**, 307–310 (2013).
- Ayarza, P. *et al.* Crustal thickness and velocity structure across the Moroccan Atlas from long offset wide-angle reflection seismic data: the SIMA experiment. *Geochim. Geophys. Res.* **15**, 1698–1717 (2014).
- Tao, W. C. & O’Connell, R. J. Ablative subduction: a two-sided alternative to convective subduction model. *J. Geophys. Res.* **97**, 8877–8904 (1992).
- Kaislaniemi, L. & van Hunen, J. Dynamics of lithospheric thinning and mantle melting by edge-driven convection: application to Moroccan Atlas mountains. *Geochim. Geophys. Res.* **15**, 3175–3189 (2014).

Acknowledgements We thank R. Govers for suggestions that improved the clarity and quality of the manuscript, and E. Engquist for aid in using the Rice DAVinCI Visualization Laboratory. We especially thank M. Harnafi and the Scientific Institute of Rabat for their contributions to the project. This research was supported by US National Science Foundation grants EAR 0003572, 0607801 and 0808939 (A.L.), EAR 0808931 (E.D.H.), EAR 0809023 and 1054638 (M.S.M.), the Venezuelan National Fund for Science, Technology and Innovation grant G-2002000478 and PDVSA-INTEVEP-FUNVISIS cooperative agreement 2004-141 (M.S.), the Spanish Ministry of Science and Innovation grants CSD2006-00041, CGL2009-09727 and CGL2010-15146 (J.G. and R.C.), and by an A. v. Humboldt Foundation Research Prize (A.L.).

Author Contributions M.S., J.G., R.C., E.D.H., F.N., M.S.M. and A.L. oversaw different aspects of the field data acquisition. A.L., F.N., S.M.T., J.M., M.S. and R.C. contributed to the receiver function data analysis. I.P., F.N., J.M., M.S.M., J.G. and A.L. contributed to the Rayleigh tomography analysis. M.J.B., E.D.H. and A.L. contributed to the body wave tomography analysis. M.S., R.C. and J.G. provided geologic, tectonic and geophysical background, which allowed A.L. and E.D.H. to pose the lithosphere removal hypothesis for testing. M.J.B., I.P., S.M.T., J.M. and A.L. constructed and interpreted the 3D images. A.L. primarily wrote the manuscript, with substantive input from E.D.H., M.J.B. and F.N. and with additional input from all of the co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.L. (alan@rice.edu).

METHODS

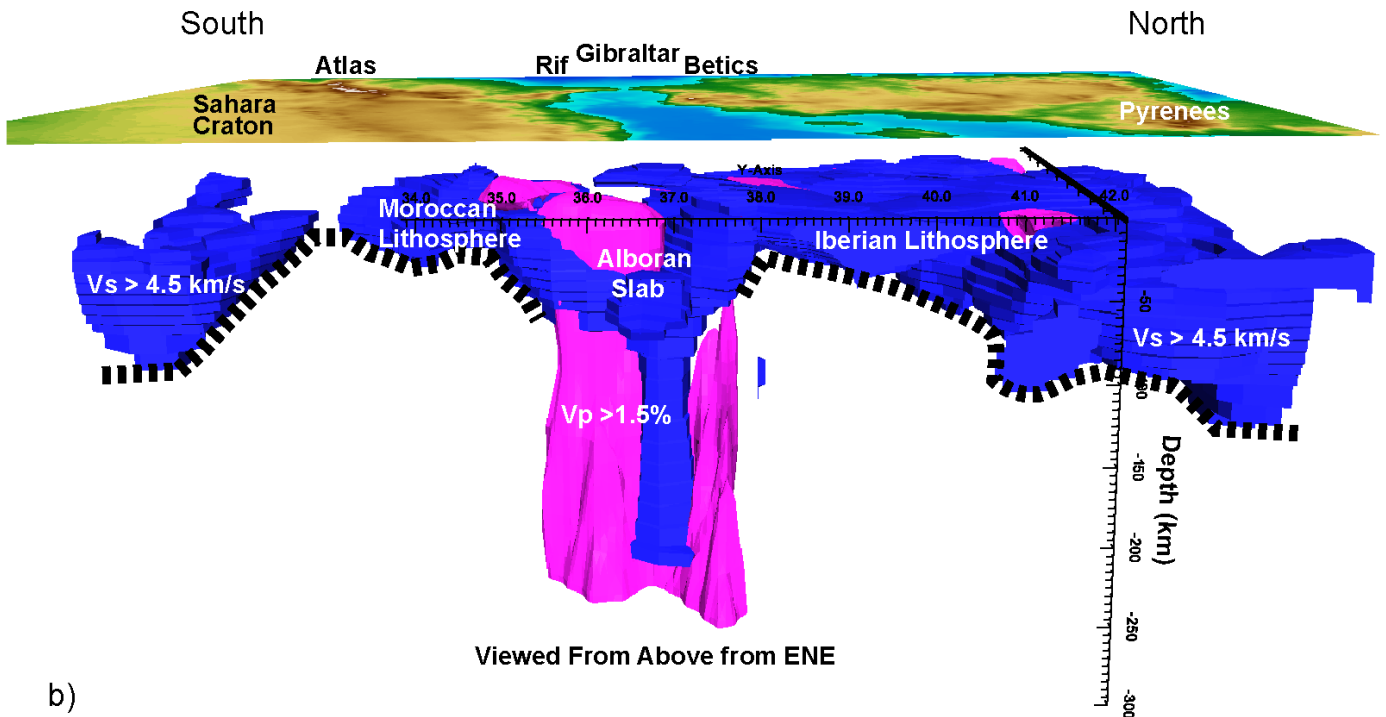
Tomography. Regional finite-frequency seismic body wave tomography is a means of including Fresnel zone effects in constructing three-dimensional Earth images from teleseismic travel time residuals. Nonlinear tomography incorporates three-dimensional ray tracing through the imaged structure⁷. The finite-frequency Rayleigh wave tomography^{34,35} images are from regional studies of the southeastern Caribbean plate boundary and the Gibraltar arc^{8,9}. This form of Rayleigh wave tomography gives absolute radially averaged V_S measurements, and, because the waves travel horizontally, provides somewhat better vertical resolution of upper mantle V_S structure than teleseismic body wave tomography. Rayleigh waves in the band we examined (0.0067–0.05 Hz) are sensitive to V_S structure from the mid-crust, ~20 km depth, into the upper mantle, ~250 km depth.

Receiver functions. Receiver function imaging makes a scattered wave image of subsurface seismic impedance boundaries using P to S or S to P converted waves from teleseismic earthquakes³⁶. Ps receiver functions are made by deconvolving direct P wave on the vertical component of motion from the radial motion to remove the earthquake source function, and replace it with a known shaping pulse. Using the incidence angle of the P wave and an estimated velocity model, the receiver functions are individually back-projected to the conversion points. The partial images made from many earthquakes recorded by many stations are summed, providing a three-dimensional image and improving signal to noise ratio³⁷. Fresnel zone effects can be included in the stacking procedure³⁸. The resulting three-dimensional image volume is a common conversion point stacked image. The process for making Sp receiver functions is similar but includes rotation of the seismograms into the direction of the S wave at the surface. Direct teleseismic S waves have lower frequency (0.2–0.05 Hz) content than the teleseismic P waves (1.0–2.0 Hz), and therefore produce a lower resolution image of the subsurface, but have other advantages, notably, Sp receiver functions are not contaminated by families of multiple S and P wave reverberations in the crust.

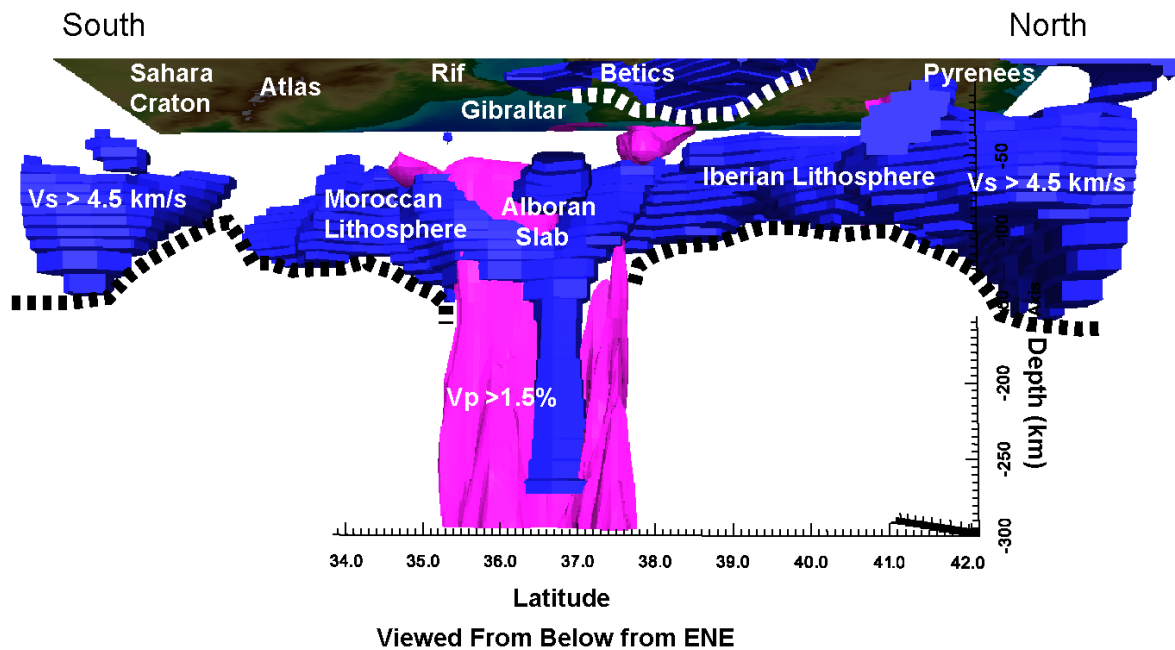
Composite seismic images. Composite seismic images are made by superimposing elements of images made from different seismic measurements. In three dimensions, volumes for superposition are constructed by choosing threshold values for isosurfaces. Since different seismic probes identify different earth structures, the isosurface thresholds are chosen to represent some aspect of seismic structure that is either unique or common to the different measurements. Regional body wave tomography is good at resolving lateral velocity variations such as those due to a subducting slab. For slabs identified in body wave tomography we chose an isosurface enclosing $\ln V_P \geq +1.5\%$. Surface wave tomography, which measures absolute velocity, is better for identifying the high and low velocities in the upper mantle lithosphere and asthenosphere, respectively, and when used in conjunction with receiver functions provides a robust means of determining lithosphere thickness. We chose an isosurface enclosing $V_S \geq 4.5 \text{ km s}^{-1}$ as indicative of the lithospheric mantle. Combined the two volumes show both lithosphere and slab structure, which neither seismic probe determines completely by itself.

34. Forsyth, D. W. & Li, A. in *Seismic Earth: Array Analysis of Broadband Seismograms* (eds Levander, A. & Nolet, G.) 81–97 (Geophys. Monogr. Ser. Vol. 157, American Geophysical Union, 2005).
35. Yang, Y. & Forsyth, D. W. Rayleigh wave phase velocities, small-scale convection and azimuthal anisotropy beneath southern California. *J. Geophys. Res.* **111**, B07306 <http://dx.doi.org/doi:10.1029/2005JB004180> (2006).
36. Rondenay, S. Upper mantle imaging with array recordings of converted and scattered teleseismic waves. *Surv. Geophys.* **30**, 377–405 (2009).
37. Dueker, K. G. & Sheehan, A. F. Mantle discontinuity structure from midpoint stacks of converted P to S waves across the Yellowstone hotspot track. *J. Geophys. Res.* **102**, 8313–8327 (1997).
38. Levander, A. & Miller, M. S. Evolutionary aspects of lithospheric discontinuity structure in the western U.S. *Geochem. Geophys. Geosyst.* **13**, Q0AK07 <http://dx.doi.org/10.1029/2012GC004056> (2012).

a)

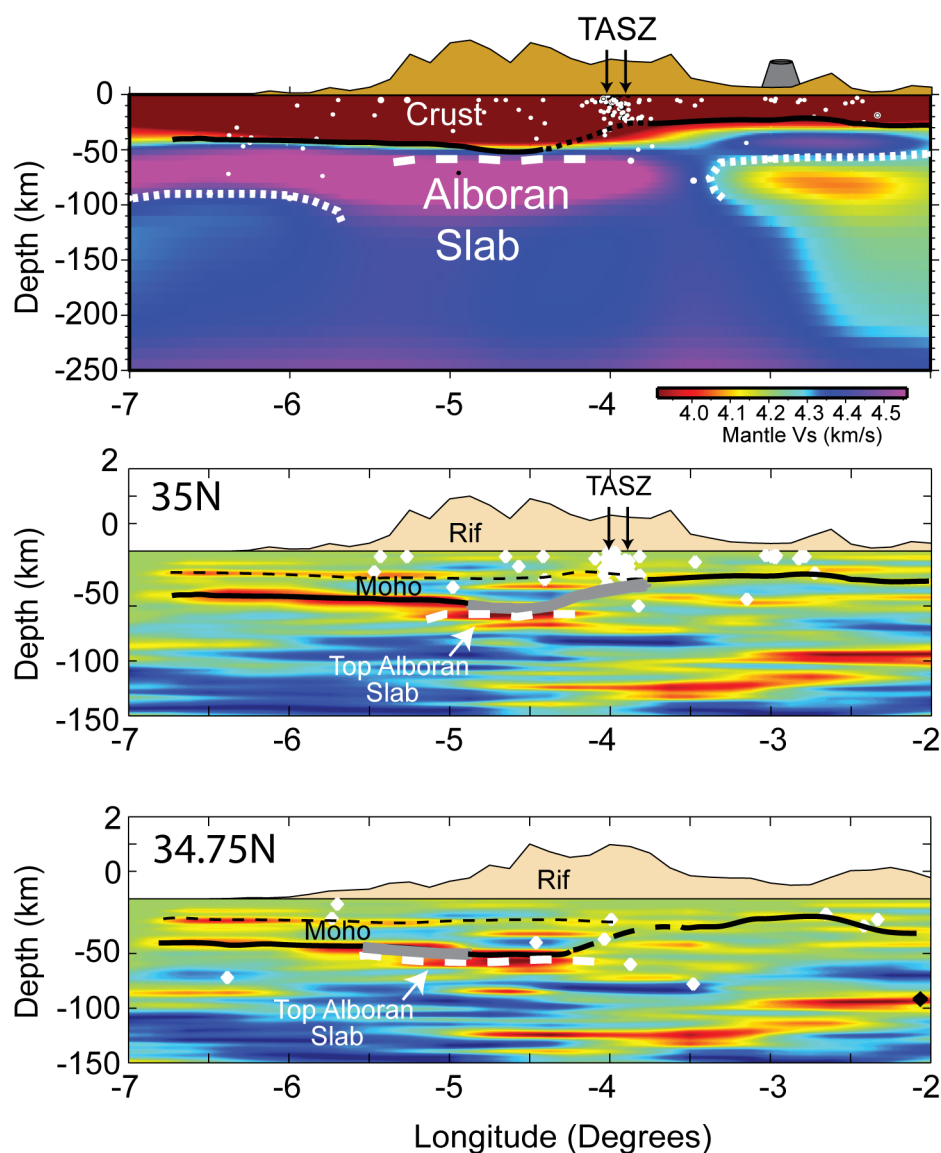


b)



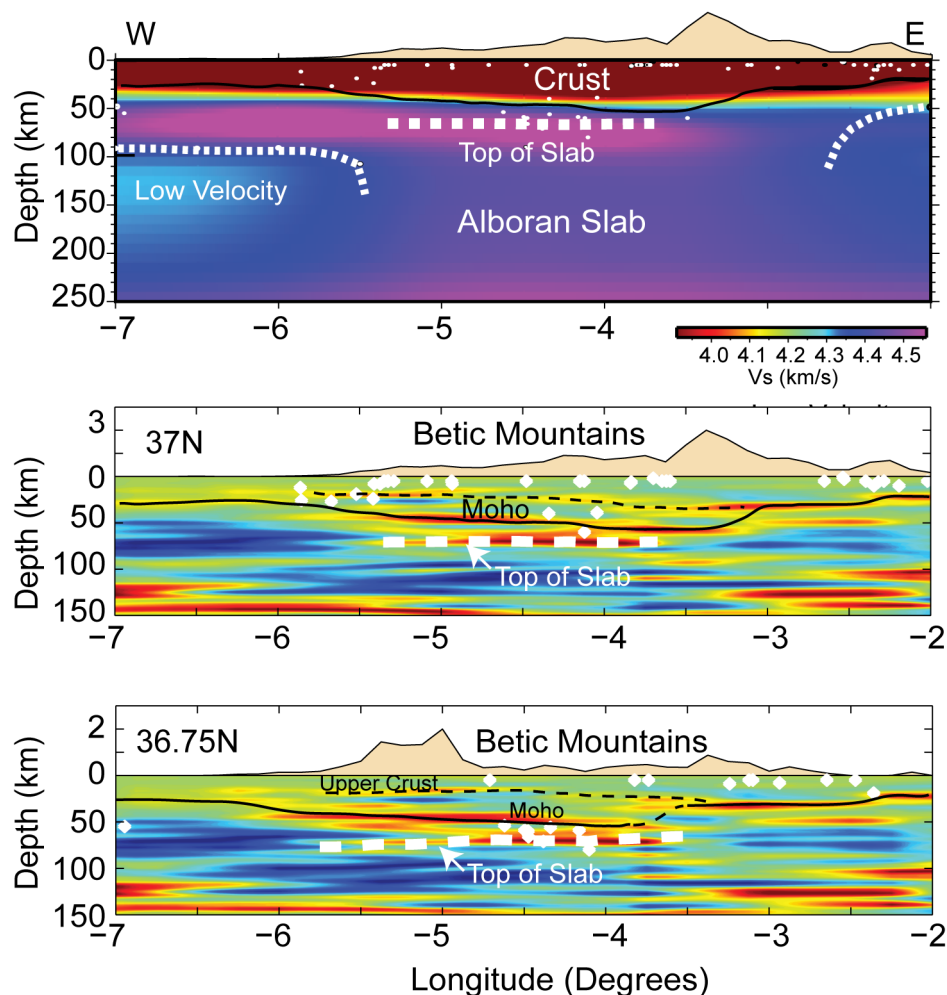
Extended Data Figure 1 | Composite seismic image showing the top of the Alboran slab and the lithosphere beneath the Gibraltar arc. **a**, Top panel is viewed from above from the east-northeast. Topography is shown at the top of the panel. The bottom of the panel is a composite of a P-body wave tomography image showing the slab (magenta, with the isosurface enclosing $\ln V_p \geq 1.5\%$), and a Rayleigh wave tomography image showing the top of the

slab and the lithosphere (blue, with the isosurface enclosing $V_s > 4.5 \text{ km s}^{-1}$). The dashed black line outlines the bottom of the lithosphere. Note that these lines do not represent depth in the perspective view. **b**, Same azimuthal view as **a** but viewed from below. The black and white dashed lines outline the bottom of the lithosphere.



Extended Data Figure 2 | Surface wave tomography model and receiver function images from northern Morocco. Top panel, Rayleigh wave tomography model along 35° N. Middle and bottom panels, 2 Hz Ps receiver function CCP stacks along 35° N (middle) and 34.75° N (bottom) showing the top of the lower crust (dashed black lines), the Moho (solid black line) and the top of the Alboran slab (dashed white line) beneath the Moroccan Rif.

In the two receiver function images the Moho and the top of the Alboran slab merge at ~50 km depth at -4.5° and diverge to the east. Moho depth from unpublished refraction profiles is shown by heavy grey line. Seismicity, shown as white diamonds, is concentrated at the Trans-Alboran shear zone (TASZ). The seismic images are shown with no vertical exaggeration.



Extended Data Figure 3 | Surface wave tomography model and receiver function images from southern Spain. Top panel, Rayleigh wave tomography model along 37° N. Middle and bottom panels, 2 Hz Ps receiver function CCP stacks along 37° N (middle) and 36.75° N (bottom) showing the top of the lower crust (dashed black lines), the Moho (black solid lines) and the top of the

Alboran slab (heavy dashed white line) beneath the Betics. In the two receiver function images the Moho and the top of the Alboran slab merge at ~50–55 km depth at ~4° and diverge in either direction. Seismicity, shown as white diamonds, occurs in the upper crust and in the zone of detachment near the base of the Iberian crust and the top of the Alboran slab.

Resolving the long-standing enigmas of a giant ornithomimosaur *Deinocheirus mirificus*

Yuong-Nam Lee¹, Rinchen Barsbold², Philip J. Currie³, Yoshitsugu Kobayashi⁴, Hang-Jae Lee¹, Pascal Godefroit⁵, François Escuillie⁶ & Tsogtbaatar Chinzorig²

The holotype of *Deinocheirus mirificus* was collected by the 1965 Polish–Mongolian Palaeontological Expedition at Altan Uul III in the southern Gobi of Mongolia¹. Because the holotype consists mostly of giant forelimbs (2.4 m in length) with scapulocoracoids², for almost 50 years *Deinocheirus* has remained one of the most mysterious dinosaurs. The mosaic of ornithomimosaur and non-ornithomimosaur characters in the holotype has made it difficult to resolve the phylogenetic status of *Deinocheirus*^{3,4}. Here we describe two new specimens of *Deinocheirus* that were discovered in the Nemegt Formation of Altan Uul IV in 2006 and Bugiin Tsav in 2009. The Bugiin Tsav specimen (MPC-D 100/127) includes a left forelimb clearly identifiable as *Deinocheirus* and is 6% longer than the holotype. The Altan Uul IV specimen (MPC-D 100/128) is approximately 74% the size of MPC-D 100/127. Cladistic analysis indicates that *Deinocheirus* is the largest member of the Ornithomimosauria; however, it has many unique skeletal features unknown in other ornithomimosaurs, indicating that *Deinocheirus* was a heavily built, non-cursorial animal with an elongate snout, a deep jaw, tall neural spines, a pygostyle, a U-shaped furcula, an expanded pelvis for strong muscle attachments, a relatively short hind limb and broad-tipped pedal unguals. Ecomorphological features in the skull, more than a thousand gastroliths, and stomach contents (fish remains) suggest that *Deinocheirus* was a megaomnivore that lived in mesic environments.

Theropoda Marsh, 1881

Ornithomimosauria Barsbold, 1983

Deinocheiridae Osmólska and Roniewicz, 1970

Deinocheiridae. *Deinocheirus mirificus* and all taxa sharing a more recent common ancestor with it than with *Ornithomimus velox*.

Revised diagnosis of the Deinocheiridae. Ornithomimosaurs in which radius and ulna well-separated; flexor tubercle of manual ungual proximally positioned; cnemial crest of tibia projects strongly anterodorsally.

Deinocheirus mirificus Osmólska and Roniewicz, 1970

Holotype. Paleontological Center of Mongolian Academy of Sciences (Ulaanbaatar, Mongolia) MPC-D 100/18 (formerly ZPal MgD-I/6) includes pectoral girdles, forelimbs, and fragments of vertebrae, ribs and gastralia².

Referred material. MPC-D 100/127, nearly complete skeleton lacking mid-dorsal vertebrae, many caudals, and right forelimb. MPC-D 100/128,

post-cervical vertebrae, ilia, partial ischia, and left hind limb (Fig. 1, Supplementary Information and Supplementary Data).

Horizon and localities. Nemegt Formation (Upper Campanian or Lower Maastrichtian) at Altan Uul III (holotype), Altan Uul IV (MPC-D 100/128), and Bugiin Tsav (MPC-D 100/127), Mongolia.

Description. Bivariate comparisons of the skull (1024 mm from premaxilla to occipital condyle) with the femur show that the cranium (Fig. 2) is low and narrow like other ornithomimosaurs, but that the antorbital region is more elongate than in other members of this clade. The premaxillae and dentaries expand anterolaterally to form a spatulate beak; the external nares open dorsally. In lateral view, the buccal edges of the premaxilla, maxilla and dentary slope anteroventrally. Pitting of the anterior surfaces indicates the presence of keratinous rhampotheca in life. The nasal is a narrow strap-like bone extending from the anterior third of the internarial bar to above the orbits. Similar to *Gallimimus*, the jugal and quadratojugal form an extensive posteroventral lobe that closed off the lower part of the infratemporal fenestra⁵. The pneumatic fossa of the quadrate is particularly large, about 30% the height of the bone. The outside diameter of the sclerotic ring (84 mm) is relatively small compared with skull length (Extended Data Fig. 1), which suggests that *Deinocheirus* was probably diurnal⁶. The edentulous lower jaw is strikingly massive and deep in lateral view in comparison with the slender, low edentulous maxilla. It scales with tyrannosaurids in the depth of the jaw, rather than ornithomimids (Extended Data Table 1 and Extended Data Fig. 2).

Each of the ten cervical vertebrae is low and long. Posteriorly the centra become progressively shorter and more trapezoidal in outline; this produces a more strongly S-curved neck (to support the relatively larger skull) than in other ornithomimosaurs (Extended Data Fig. 3). Twelve dorsal vertebrae have distinctive configurations; the neural spines of anterior dorsals are low, but increase progressively to that of the last dorsal, which is 8.5 times taller than its centrum height; it almost attains the highest ratio of *Spinosaurus*⁷. Because of the presence of an intricate system of interspinous ligaments, they were probably specialized to support the abdomen from the hips and hind limbs in a manner similar to an asymmetrical cable-stayed bridge. The elongate spines might also have served in display. Dorsal vertebrae have numerous laminae forming basal webbing with associated fossae, and the degree of pneumaticity is comparable with that of a sauropod (Extended Data Fig. 4).

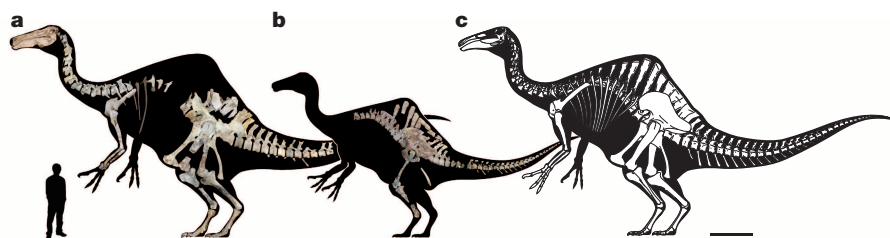


Figure 1 | *Deinocheirus mirificus*. a, MPC-D 100/127. b, MPC-D 100/128. c, Composite reconstruction of MPC-D 100/127 with a simple proportional enlargement of MPC-D 100/128. Scale bar, 1 m. The human outline is 1.7 m tall. The holotype and the two new specimens provide almost all skeletal information of *Deinocheirus*.

¹Geological Museum, Korea Institute of Geoscience and Mineral Resources, Daejeon 305-350, South Korea. ²Paleontological Center, Mongolian Academy of Sciences, Ulaanbaatar 210-351, Mongolia. ³Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada. ⁴Hokkaido University Museum, Hokkaido University, Sapporo 060-0810, Japan. ⁵Earth and History of Life, Royal Belgian Institute of Natural Sciences, Rue Vautier 29, 1000 Bruxelles, Belgium. ⁶Eldonia, 9 Avenue des Portes Occitanes, 3800 Gannat, France.

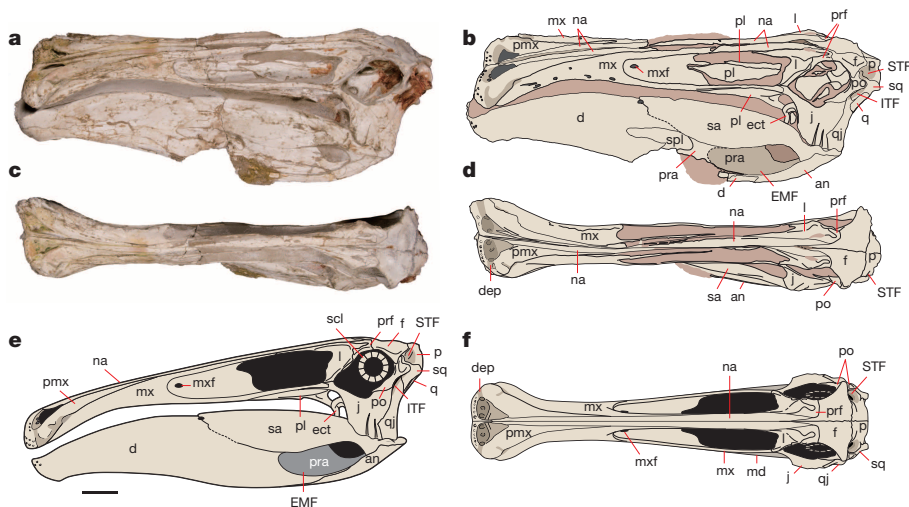


Figure 2 | Skull of *Deinocheirus mirificus* (MPC-D 100/127). **a, b**, Left lateral view. **c, d**, Dorsal view. **e**, Reconstruction in composite lateral view. **f**, Reconstruction in dorsal view. Scale bar, 10 cm. Abbreviations: an, angular; d, dentary; dep, depression; ect, ectopterygoid; EMF, external mandibular fenestra; f, frontal; ITF, infratemporal fenestra; j, jugal; l, lacrimal; md, mandible; mx, maxilla; mxf, maxillary fenestra; na, nasal; p, parietal; pl, palatine; pmx, premaxilla; po, postorbital; pra, prearticular; prf, prefrontal; q, quadrate; qj, quadratojugal; sa, surangular; scl, sclerotic ring; spl, splenial; sq, squamosal; STF, supratemporal fenestra.

All six sacral neural spines are tall and highly pneumatic, extending up to 170% of ilium height. Except for the first sacral, the tops of the sacral neural spines are fused into a midline plate of bone, the dorsal margin of which is straight in lateral view. An accessory spinodiapophyseal lamina is developed anterior to the spinodiapophyseal lamina in each of the third to fifth sacra. These two laminae meet at the diapophysis to form a unique V-shape in lateral view (Fig. 3).

The first caudal neural spine curves anteriorly in lateral view, and is distinctly shorter than that of the final sacral. The tall anterior caudal neural spines have strong anterior and posterior rugosities with many small foramina for the interarcual ligaments. The end of the tail is

represented by at least two vertebrae that were fused together as in oviraptorosaur and therizinosauroid pygostyles^{8,9}. The presence of a pygostyle suggests the possibility that ornithomimosaurs, which are known to have pennaceous feathers¹⁰, also had fans of feathers at the ends of their tails for display purposes¹¹. Honeycombed camellate pneumatic systems are present in all *Deinocheirus* vertebrae except for the atlas and the distal caudals. Camellate internal structure is also in the parts of the ilium, pubis and ischium that are adjacent to the sacra. Vertebral pneumaticity correlates with gigantism¹².

A U-shaped furcula with a hypocleidium was recovered for the first time in ornithomimosaurs. The triangular acromial process of the

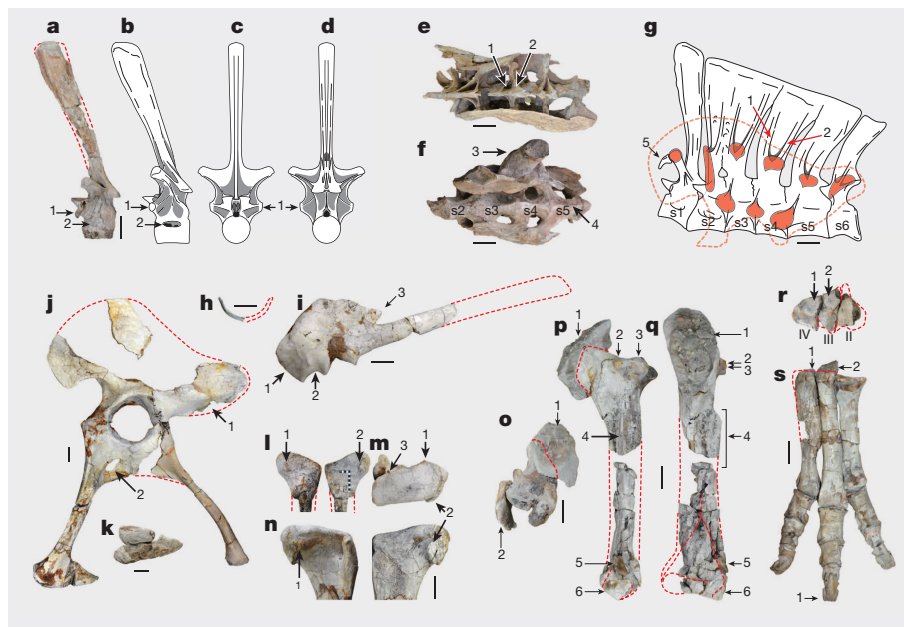


Figure 3 | Postcranial skeletons of *Deinocheirus mirificus* (MPC-D 100/127, MPC-D 100/128). **a–d**, Dorsal vertebra 12 of MPC-D 100/128 (**a**, left lateral view; **b–d**, reconstruction in lateral, anterior, and posterior views; shaded areas indicate fossae; 1, parapophysis; 2, pneumatopore; see Supplementary Information for detailed skeletal anatomy). **e–g**, Sacrum with ilia of MPC-D 100/128 (**e**, dorsal view; **f**, ventral view; **g**, reconstruction of sacrum; orange shading indicates contacts for ilium; 1, accessory spinodiapophyseal lamina; 2, spinodiapophyseal lamina; 3, left proximal femur; 4, sharp ventral keel; 5, diminutive sacral 1 rib). **h**, Furcula of MPC-D 100/127. **i**, Left scapulocoracoid of MPC-D 100/127 (1, subquadrangular coracoid with ventrally extended blade; 2, unexpanded subglenoid fossa; 3, acromial process). **j, k**, Pelvic girdle of MPC-D 100/127 (**j**, lateral view; 1, anteroventrally

inclined brevis fossa; 2, completely enclosed pubic obturator foramen; **k**, pubic boot in distal view). **l–n**, Left femoral heads of MPC-D 100/128 and MPC-D 100/127 (**l**, anterior (left) and posterior (right) views of MPC-D 100/128; **m**, proximal view of MPC-D 100/127; **n**, anterior (left) and posterior (right) views of MPC-D 100/127; 1, dome-like tubercle; 2, pronounced vertical hook; 3, intertrochanteric foramen). **o–q**, Left tibia of MPC-D 100/127 with gastrolith mass *in situ* (**o**, proximal view; **p**, lateral view; **q**, anterior view; 1, gastrolith mass; 2, accessory process; 3, lateral condyle; 4, fibular flange; 5, distal part of fibula; 6, calcaneum). **r, s**, Foot of MPC-D 100/127 (**r**, proximal view of right metatarsal (left–right reversal); 1 and 2, tarsals; **s**, dorsal view of left foot; 1, blunt tip of ungual). All scale bars, 10 cm.

scapulocoracoid prominently projects at an angle of about 50° from the scapular blade. This feature cannot be seen in the holotype due to damage. The ilium appears distinctive because regions of the typical ornithomimosaur pelvis are hypertrophied to support the animal's great weight. The rounded preacetabular ala is taller but anteroposteriorly shorter than the postacetabular ala, which has a concave dorsal margin. The robust supraacetabular crest overhangs the anterodorsal margin of the acetabulum; because the crest was normally directly above the head of the femur, the ilium was tilted posteroventrally in life. The ischium is as long as the pubis.

The femur is longer than the tibia, as expected for such a large animal¹³. In dorsal view, the medial edge of the femoral head has a prominent hook on the posteromedial corner, and a dome-like tubercle on the anterior edge that has never been reported, to our knowledge, in any other dinosaur. The femoral head is twisted 15° anteromedially to the femoral shaft. The robust cnemial crest of the tibia projects above the level of the proximal articular surface as in *Beishanlong*¹⁴ and *Garudimimus*¹⁵. The baseline condition of a relatively short, non-arctometatarsalian metatarsus is present as in most other theropods¹⁶. However, the outline of the proximal end of the third metatarsal in *Deinocheirus* is quadrangular, rather than triangular as in basal ornithomimosaur. The foot is as short in the relative lengths of digit III to metatarsal III as it is in *Garudimimus*. Each pedal ungual has a unique shape that has never been reported, to our knowledge, in any other theropod dinosaur, with the distal end bluntly truncated rather than tapered (Extended Data Fig. 5).

Deinocheirus was coded into a theropod data matrix that incorporated a comprehensive analysis for Ornithomimosauria¹⁷ (Extended Data Fig. 6). It was recovered as a sister taxon of *Garudimimus* within the Ornithomimosauria (Fig. 4). This comprehensive phylogeny suggests that basal ornithomimosaur separated into two lineages (Deinocheiridae and Ornithomimidae) in the Early Cretaceous. The Deinocheiridae consists of *Deinocheirus*, *Garudimimus*, and the older *Beishanlong*. *Deinocheirus* (MPC-D 100/127) had a body length of 11 m and an estimated body weight of 6,358 kg¹⁸, and is the largest known ornithomimosaur. Deinocheirids followed a different evolutionary path from cursorial ornithomimids. Limb proportions suggest that, like *Deinocheirus*, *Garudimimus* was not adapted for speed. Deinocheirids also lack the distinct insertion of the iliofibularis muscle onto the fibular shaft, and in lateral view do not have the straight pedal unguals seen in other ornithomimosaur.

A large number (>1,400) of gastroliths (ranging from 8 to 87 mm) were collected during excavation from inside the ribs and gastralia of MPC-D 100/127. The ratio of gastrolith mass/body mass (approximately 0.0022)¹⁹ suggests that gastroliths were used to grind food in the toothless *Deinocheirus*. Rhamphotheca, edentulous jaws, a U-shaped dentary symphysis, an anteriorly downturned dentary, and a dorsally convex dentary in *Deinocheirus* are known as characters of at least facultative herbivory in Coelurosauria^{20,21}. Interestingly, fish remains (vertebrae and scales) were discovered *in situ* amongst the gastroliths (Extended Data Fig. 7). The stomach contents and various herbivorous ecomorphological features suggest that *Deinocheirus* was omnivorous.

The supratemporal and infratemporal adductor chambers in *Deinocheirus* are strikingly small compared with the elongate, large skull, implying weak bite force. The unique *Deinocheirus* skull morphology and a simple orthal jaw action is more suitable for cropping relatively soft understory vegetation (or possibly herbaceous water plants): skull length correlates negatively with bite force and feeding height (plant height)²². The deep buccal cavity, as evidenced by the deep lower jaws, suggests the presence of a massive tongue that when manipulated would create suction for ingesting the organic material cropped and disturbed by the broad bill as it foraged on the bottom of streams, lakes and ponds. The duck-like bill of *Deinocheirus* may be ecologically tied to water-based food as in ducks²³, as well as ground-level, non-selective browsing as in some genera of sauropods²⁴ and hadrosaurs²⁵.

Stomach contents include fish remains and indicate that *Deinocheirus* frequented freshwater habitats. The short and compact manual claws of *Deinocheirus* are similar morphologically to those of the therizinosaur

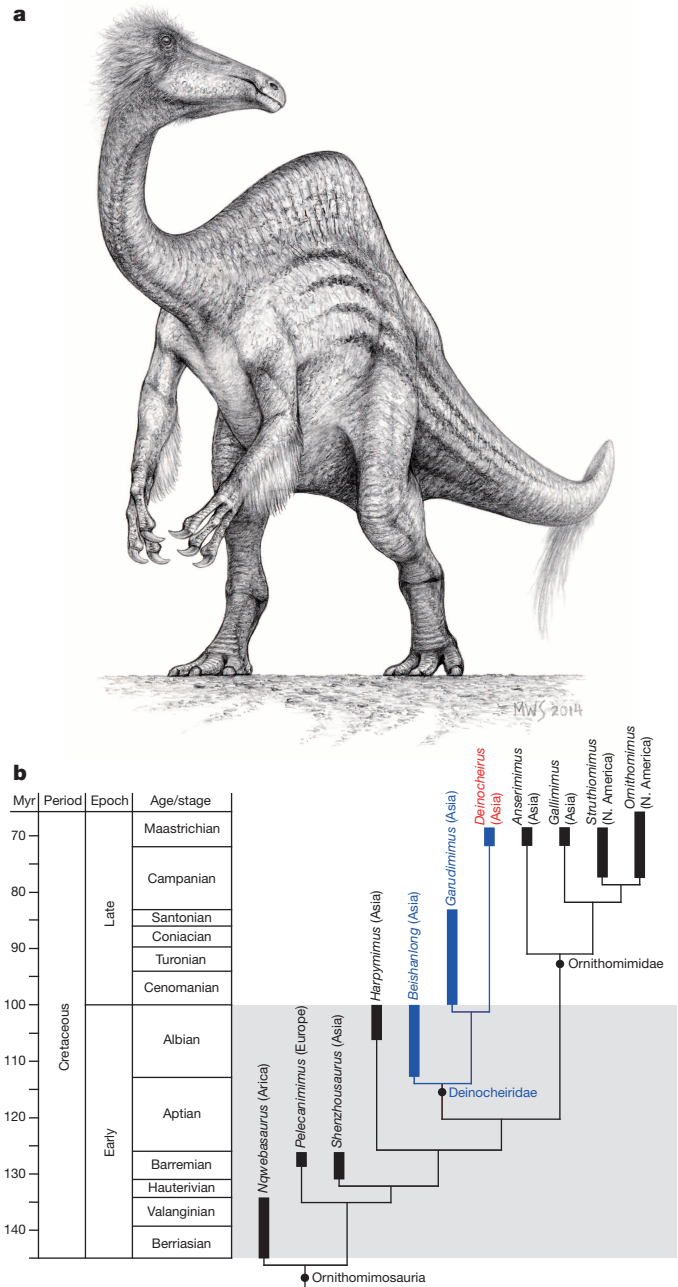


Figure 4 | Phylogenetic relationships of *Deinocheirus mirificus* within Ornithomimosauria. **a**, Hypothetical fleshed-out reconstruction of *Deinocheirus mirificus* (by Michael Skrepnick). **b**, Time-scaled strict consensus tree of the six most-parsimonious trees from our analysis (tree length = 2,927, consistency index = 0.22, retention index = 0.59; Supplementary Information). In this hypothesis *Deinocheirus* is a derived taxon of the Deinocheiridae, which is the sister-group of the Ornithomimidae.

*Alxasaurus*²⁶, which were used in a generalist fashion²⁷. If true, long forearms with giant claws may have been used for digging and gathering herbaceous plants. Robust hind limbs with posteroventrally tilted, wide hips, femora longer than tibiae, and massive feet clearly indicate that *Deinocheirus* was a slow mover. The blunt tips of the pedal unguals would have prevented its feet from sinking deep into wet substrates. The braided or meandering river systems of the Nemegt Formation²⁸ provided a good niche for omnivorous *Deinocheirus* to flourish.

Gigantism may well have been the way *Deinocheirus* escaped predation from *Tarbosaurus* and other theropods of the Nemegt fauna. The trade-off is that it lost the cursorial abilities of its relatives. Ornithomimosaur is normally small to medium-bodied dinosaurs; extending

the allometric trajectories of all known specimens of this clade to gigantic sizes seems to account for many of the peculiar features of *Deinocheirus*, including the elongate forelimbs. The long skull with its broad bill and deep lower jaws demonstrates a more specialized diet than other known ornithomimosaurs. The massive increase in body weight is no doubt responsible for other features, including the increase in neural spine height, and the broad-tipped pedal unguals. Additionally, *Deinocheirus* has at least two features (furculum, pygostyle) previously unknown in ornithomimosaurs, but found in a broad spectrum of other theropods. The discovery of the original specimen almost half a century ago suggested that this was an unusual dinosaur, but did not prepare us for how distinctive *Deinocheirus* is—a true cautionary tale in predicting body forms from partial skeletons, even for animals in which the relationships are known.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 August; accepted 19 September 2014.

Published online 22 October 2014.

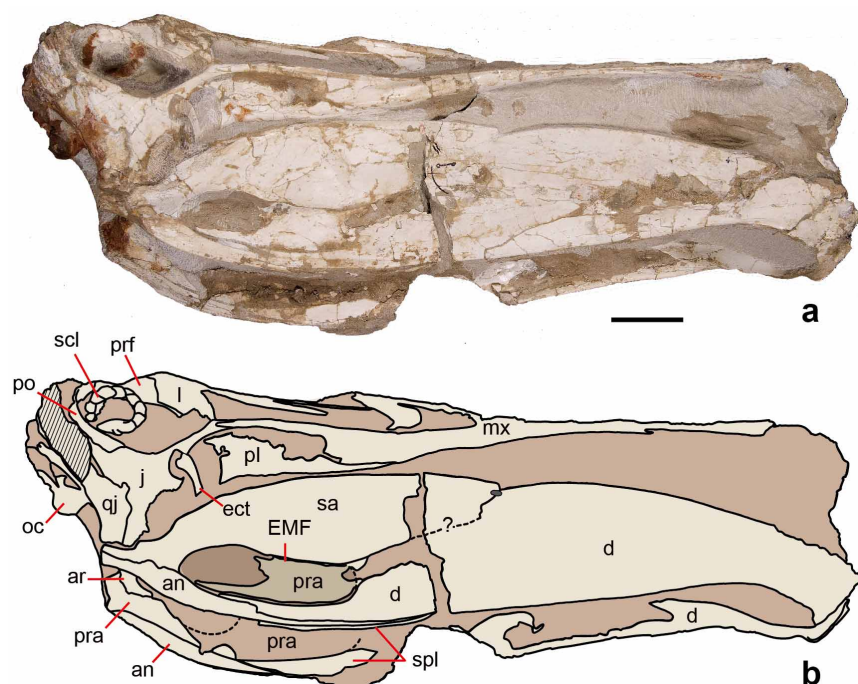
- Kielan-Jaworowska, Z. Third (1965) Polish–Mongolian Palaeontological Expedition to the Gobi Desert and western Mongolia. *Bull. de l'Acad. Pol. Sci. Cl. II* **14**, 249–252 (1966).
- Osmólska, H. & Roniewicz, E. Deinocheiridae, a new family of theropod dinosaurs. *Palaeontol. Polonica* **21**, 5–19 (1970).
- Makovicky, P. J., Kobayashi, Y. & Currie, P. J. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmólska, H.) 137–150 (Univ. California Press, 2004).
- Kobayashi, Y. & Barsbold, R. Ornithomimids from the Nemegt Formation of Mongolia. *J. Paleontol. Soc. Korea* **22**, 195–207 (2006).
- Osmólska, H., Roniewicz, E. & Barsbold, R. A new dinosaur *Gallimimus bullatus* n. gen., n. sp. (Ornithomimidae) from the Upper Cretaceous of Mongolia. *Palaeontol. Polonica* **27**, 103–143 (1972).
- Schmitz, L. & Motani, R. Nocturnality in dinosaurs inferred from scleral ring and orbit morphology. *Science* **332**, 705–708 (2011).
- Bailey, J. B. Neural spine elongation in dinosaurs: sailbacks or buffalo-backs? *J. Paleontol.* **71**, 1124–1146 (1997).
- Barsbold, R. et al. A pygostyle from a non-avian theropod. *Nature* **403**, 155–156 (2000).
- Xu, X., Cheng, Y., Wang, X. & Chang, C. Pygostyle-like structure from *Beipiaosaurus* (Theropoda, Therizinosauroidea) from the Lower Cretaceous Yixian Formation of Liaoning, China. *Acta Geol. Sin.* **77**, 294–298 (2003).
- Zelenitsky, D. K. et al. Feathered non-avian dinosaurs from North America provide insight into wing origins. *Science* **338**, 510–514 (2012).
- Persons, W. S. IV, Currie, P. J. & Norell, M. A. Oviraptorid tail forms and functions. *Acta Palaeontol. Pol.* **59**, 553–567 (2014).
- Schwarz-Wings, D. et al. Mechanical implications of pneumatic neck vertebrae in sauropod dinosaurs. *Proc. R. Soc. B* **277**, 11–17 (2010).
- Currie, P. J. Allometric growth in tyrannosaurids (Dinosauria: Theropoda) from the Upper Cretaceous of North America and Asia. *Can. J. Earth Sci.* **40**, 651–665 (2003).
- Makovicky, P. J. et al. A giant ornithomimosaur from the Early Cretaceous of China. *Proc. R. Soc. B* **277**, 191–198 (2010).
- Kobayashi, Y. & Barsbold, R. Reexamination of a primitive ornithomimosaur, *Garudimimus brevipes* Barsbold, 1981 (Dinosauria: Theropoda), from the Late Cretaceous of Mongolia. *Can. J. Earth Sci.* **42**, 1501–1521 (2005).
- Holtz, T. R., Jr. The arctometatarsalian pes, an unusual structure of the metatarsus of Cretaceous Theropoda (Dinosauria: Saurischia). *J. Vertebr. Paleontol.* **14**, 480–519 (1995).
- Choiniere, J. N., Forster, C. A. & de Klerk, W. J. New information on *Nqwebasaurus thwazi*, a coelurosaurian theropod from the Early Cretaceous Kirkwood Formation in South Africa. *J. Afr. Earth Sci.* **71–72**, 1–17 (2012).
- Benson, R. B. J. et al. Rates of dinosaur mass body evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. *PLoS Biol.* **12**, e1001853 (2014).
- Wings, O. & Sander, P. M. No gastric mill in sauropod dinosaurs: new evidence from analysis of gastrolith mass and function in ostriches. *Proc. R. Soc. B* **274**, 635–640 (2007).
- Zanno, L. E. & Makovicky, P. J. Herbivorous ecomorphology and specialization patterns in theropod dinosaur evolution. *Proc. Natl Acad. Sci. USA* **108**, 232–237 (2011).
- Zanno, L. E., Gillette, D. D., Albright, L. B. & Titus, A. L. A new North American therizinosauroid and the role of herbivory in 'predatory' dinosaur evolution. *Proc. R. Soc. B* **276**, 3505–3511 (2009).
- Mallon, J. C. & Anderson, J. S. Skull ecomorphology of magaherbivorous dinosaurs from the Dinosaur Park Formation (Upper Campanian) of Alberta, Canada. *PLoS ONE* **8**, e67182 (2013).
- Norell, M. A., Makovicky, P. J. & Currie, P. J. The beaks of ostrich dinosaurs. *Nature* **412**, 873–874 (2001).
- Whitlock, J. A. Inferences of diplodocid (Sauropoda: Dinosauria) feeding behavior from snout shape and microwear analyses. *PLoS ONE* **6**, e18304 (2011).
- Carrano, M. T., Janis, C. M. & Sepkoski, J. J. Jr. Hadrosaurs as ungulate parallels: lost lifestyles and deficient data. *Acta Palaeontol. Pol.* **44**, 237–261 (1999).
- Russell, D. A. & Dong, Z. The affinities of a new theropod from the Alxa Desert, Inner Mongolia, People's Republic of China. *Can. J. Earth Sci.* **30**, 2107–2127 (1993).
- Lautenschlager, S. Morphological and functional diversity in therizinosaur claws and the implications for theropod claw evolution. *Proc. R. Soc. B* **281**, 20140497 (2014).
- Jerzykiewicz, T. & Russell, D. A. Late Mesozoic stratigraphy and vertebrates of the Gobi Basin. *Cretac. Res.* **12**, 345–377 (1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements Information gained from Zofia Kielan-Jaworowska and Wojciech Skarzynski (who respectively found and excavated the holotype) allowed us to refine the original quarry at Altan Uul III. Thanks go to all members of Korea-Mongolia International Dinosaur Expedition (KID) in 2006 and 2009. The KID expedition was supported by a grant to Y.-N.L. from Hwaseong City, Gyeonggi Province, South Korea. Research support was from Korea Institute of Geosciences and Mineral Resources, Korea and Paleontological Center of Mongolian Academy of Sciences, Mongolia.

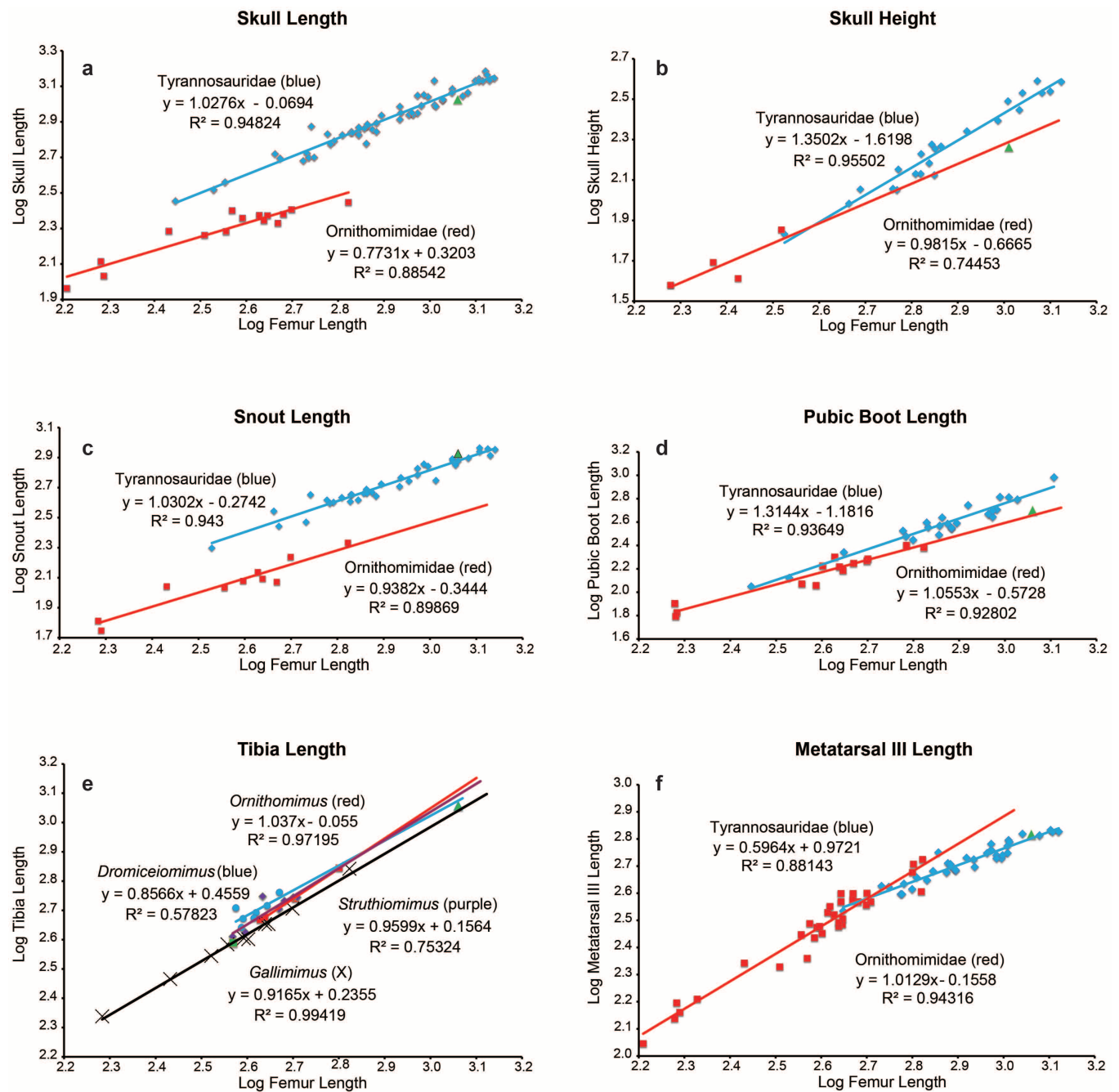
Author Contributions Y.-N.L. designed the project; Y.-N.L., R.B., P.J.C., Y.K. and H.-J.L. collected fossils and performed the research; P.J.C., P.G., F.E. and T.C. helped to repatriate the poached parts of the specimen so that they could be studied. H.-J.L. assembled figures; Y.-N.L. developed and wrote the manuscript with contributions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.-N.L. (ylee@kigam.re.kr).

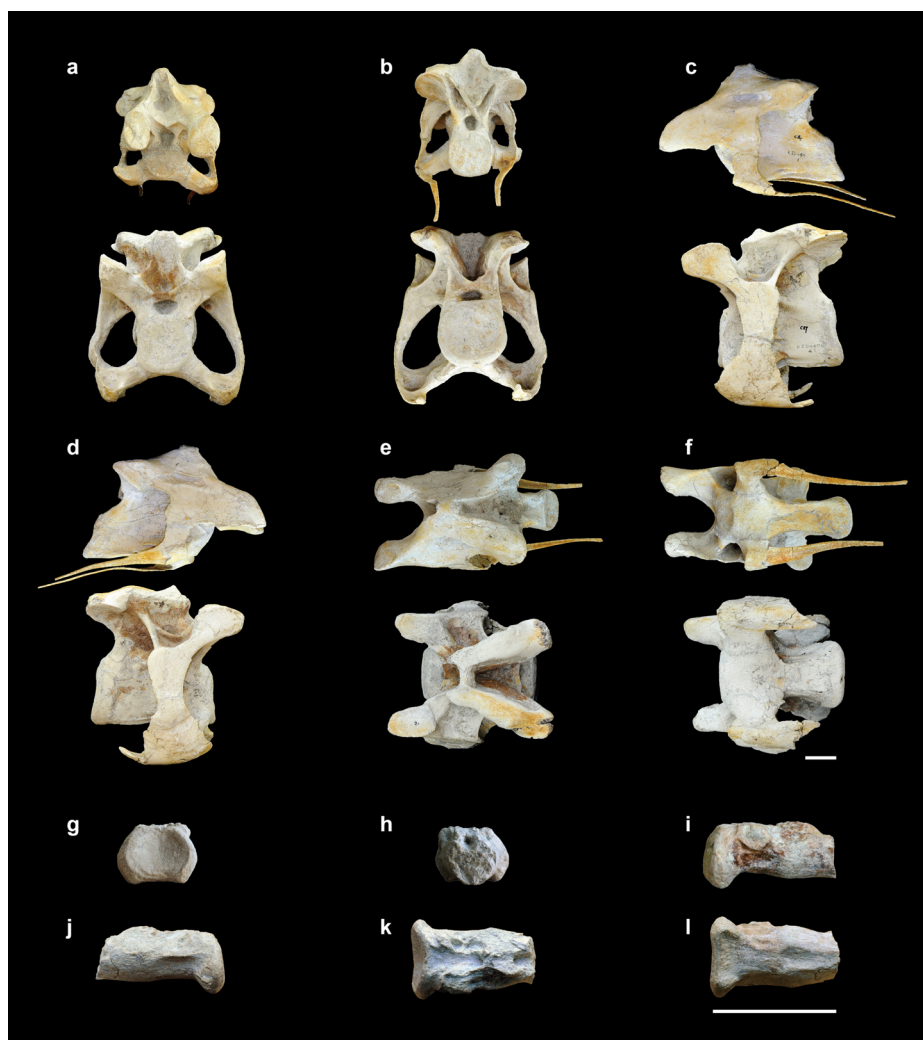


Extended Data Figure 1 | Skull of *Deinocheirus mirificus* (MPC-D 100/127). **a**, In right lateral view. **b**, Line drawing in right lateral view. Scale bar, 10 cm. Abbreviations: an, angular; ar, articular; d, dentary; ect,

ectopterygoid; EMF, external mandibular fenestra; j, jugal; l, lacrimal; mx, maxilla; oc, occipital condyle; pl, palatine; po, postorbital; pra, prearticular; prf, prefrontal; qj, quadratojugal; sa, surangular; scl, sclerotic ring; spl, splenial.

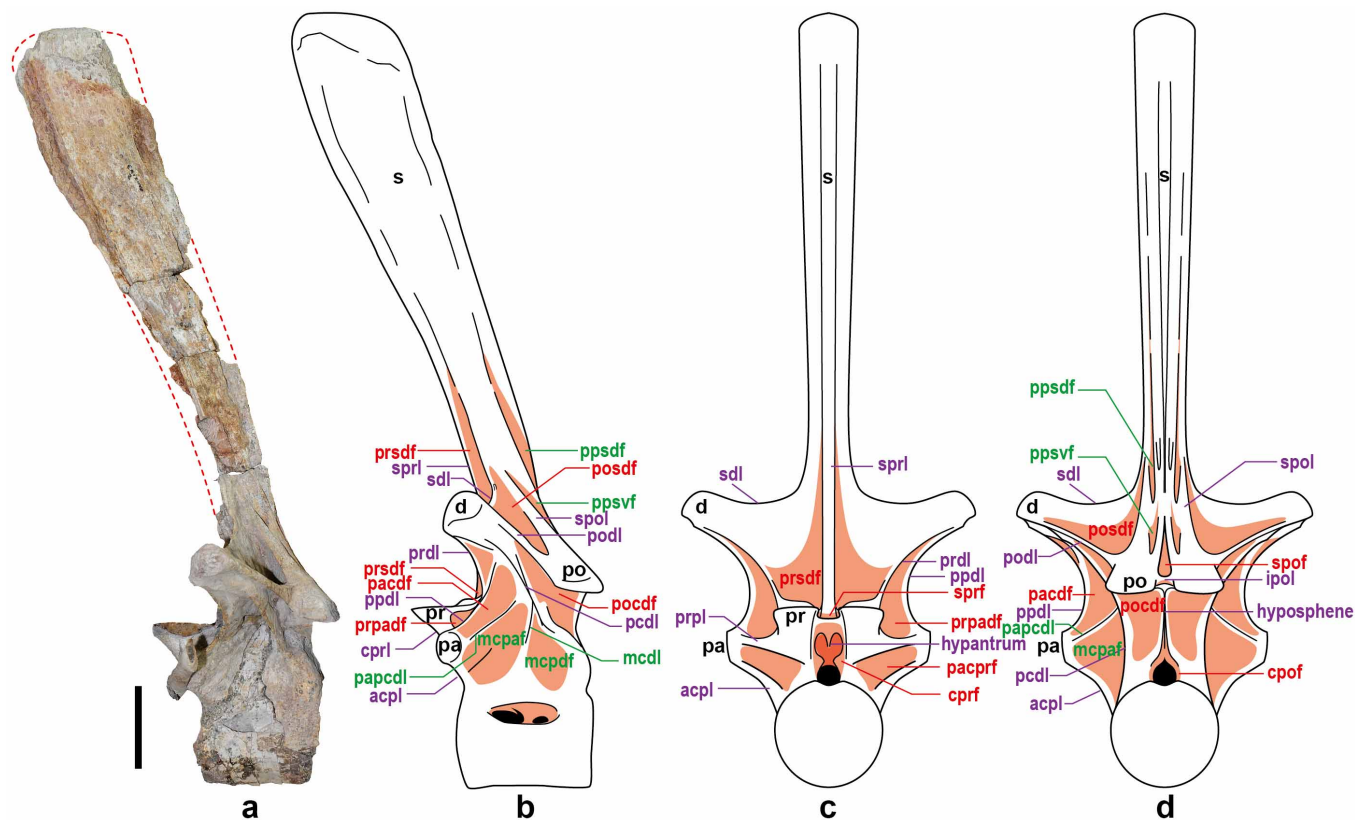


Extended Data Figure 2 | Comparisons of femur length. a, Skull length. b, Skull height. c, Snout length. d, Pubic boot length. e, Tibia length. f, Metatarsal III length in Tyrannosauridae and Ornithomimidae. The green triangle is *Deinocheirus* and the green square is *Garudimimus*.



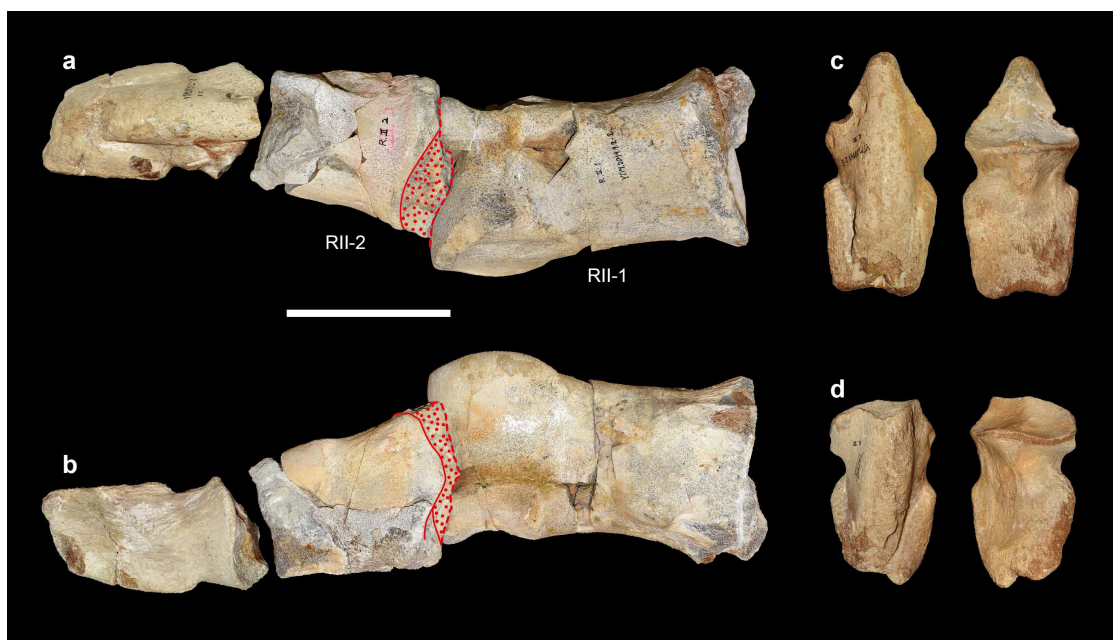
Extended Data Figure 3 | Fourth (upper) and seventh (lower) cervical vertebrae (a–f), and pygostyle (g–l) of *Deinocheirus mirificus* (MPC-D 100/127). a, In anterior view. b, In posterior view. c, In left lateral view. d, In

right lateral view. e, In dorsal view. f, In ventral view. g, In anterior view. h, In posterior view. i, In left lateral view. j, In right lateral view. k, In dorsal view. l, In ventral view. Scale bars, 5 cm.



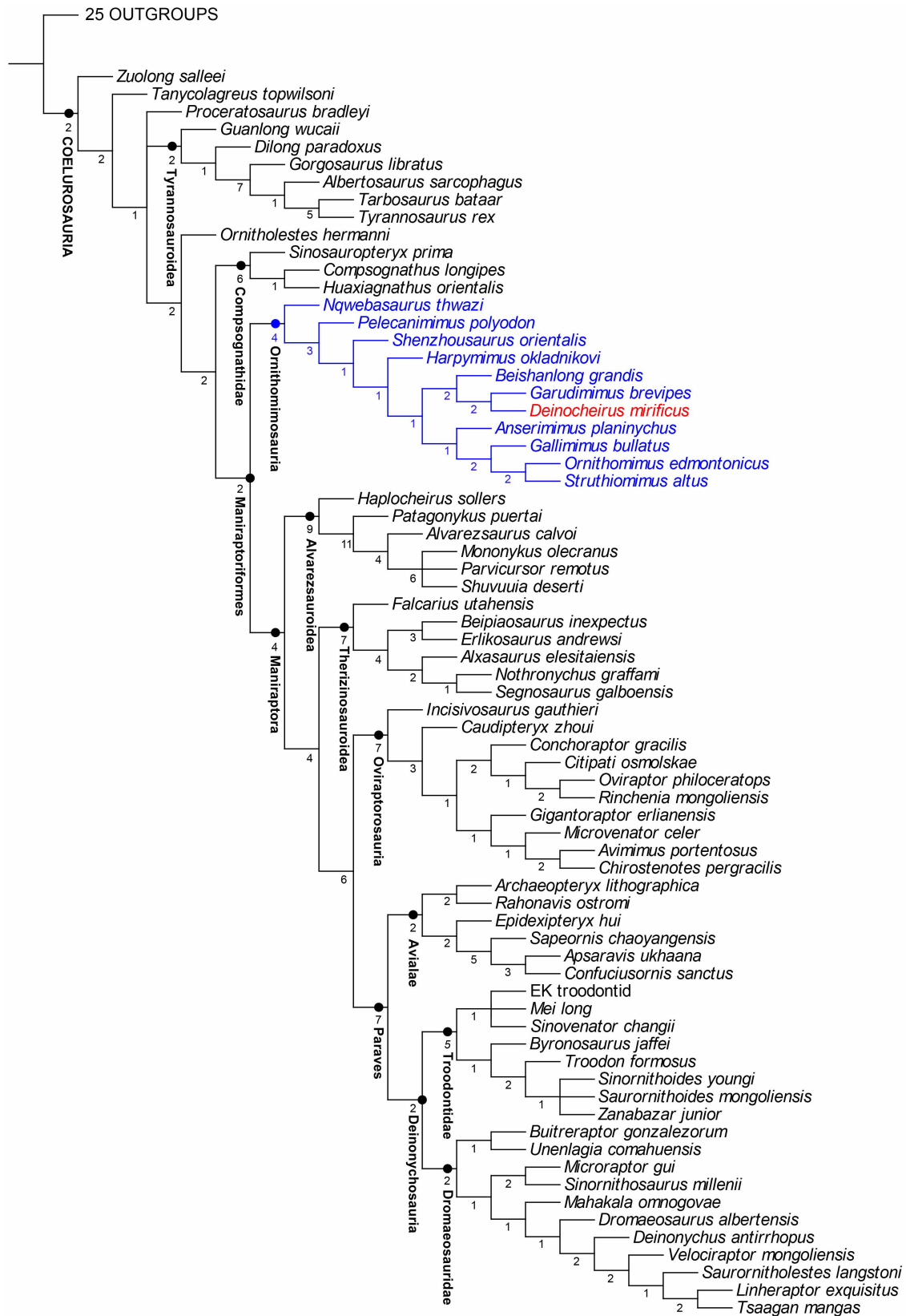
Extended Data Figure 4 | Dorsal vertebra 12 of *Deinocheirus mirificus* (MPC-D 100/128). **a**, Specimen in left lateral view. **b–d**, Reconstructions (**b**, in left lateral view; **c**, in anterior view; **d**, in posterior view). Purple and red colours indicate lamina and fossae, respectively. Green colours indicate new terms used for *Deinocheirus*. Scale bar, 10 cm. Abbreviations: acpl, anterior centroparapophyseal lamina; cpor, centropostzygapophyseal fossa; cprf, centroprezygapophyseal fossa; cprl, centroprezygapophyseal lamina; d, diapophysis; ipol, infrapostzygapophyseal lamina; mcdl, middle centrodiapophyseal lamina; mcprf, middle centroparapophyseal fossa; mcpdl, middle centropostdiapophyseal fossa; pa, parapophysis; pacdf, parapophyseal centrodiapophyseal fossa; pacprf, parapophyseal centroprezygapophyseal fossa; papcdl, parapophyseal posterior centrodiapophyseal lamina; pccl,

posterior centrodiapophyseal lamina; po, postzygapophysis; pocdf, postzygapophyseal centrodiapophyseal fossa; podl, postzygodiapophyseal lamina; posdf, postzygapophyseal spinodiapophyseal fossa; pr, prezygapophysis; ppdl, paradiapophyseal lamina; ppsdf, posterior postzygapophyseal spinodorsal fossa; ppsvf, posterior postzygapophyseal spinovertebral fossa; prdl, prezygodiapophyseal lamina; presdf, prezygapophyseal spinodiapophyseal fossa; prpadf, prezygapophyseal paradiapophyseal fossa; prpl, prezygoparapophyseal lamina; prsdf, prezygapophyseal paradiapophyseal fossa; s, neural spine; sdl, spinodiapophyseal lamina; spof, spinopostzygapophyseal fossa; spol, spinopostzygapophyseal lamina; sprf, spinoprezygapophyseal fossa; sprl, spinoprezygapophyseal lamina.

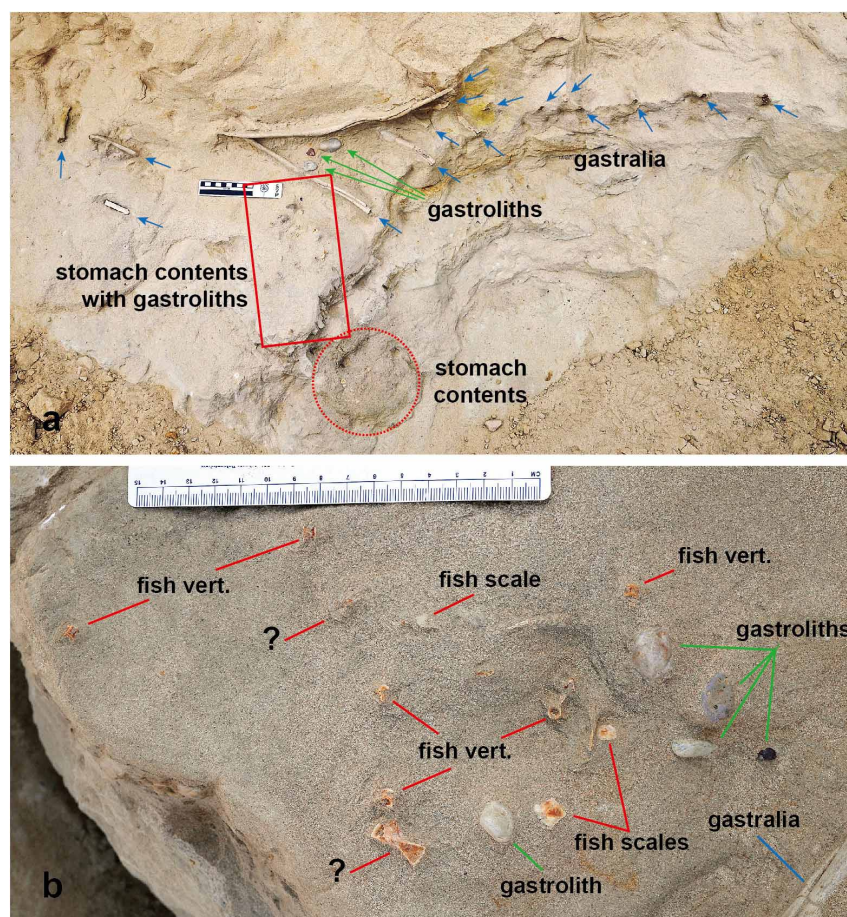


Extended Data Figure 5 | Right pedal digit II and unguals of *Deinocheirus mirificus* (MPC-D 100/127). **a, b,** Digit II in dorsal and ventral views. Isolated right pedal phalanx II-2 was collected in 2009 and perfectly fits into the impression in the matrix attached to the poached phalanx II-1 (the red dot and

solid lines indicate the contact face of the bone boundary with matrix, and the dotted area is matrix). **c,** Ungual of digit III in dorsal and ventral views. **d,** Ungual of digit IV in dorsal and ventral views. Scale bar, 10 cm.



Extended Data Figure 6 | Strict consensus topology of six most parsimonious reconstructions of theropod relationships found in the phylogenetic analysis of 568 characters and 96 taxa. Values indicate Bremer support derived from the BREMER.RUN script supplied by TNT.



Extended Data Figure 7 | Stomach contents of *Deinocheirus mirificus* (MPC-D 100/127). **a**, Photo to show *in situ* gastralia, gastroliths, and stomach contents. Blue and green arrows represent gastralia and gastroliths.

Red rectangle is an area of scattered fish remains and gastroliths. Red circle is an area where broken fish bones are aggregated. **b**, Enlarged photo of scattered fish remains (vertebrae, scales) with gastroliths in **a**.

Extended Data Table 1 | Allometric comparisons between ornithomimosaurids (or ornithomimids) and tyrannosaurids derived from the power equation $y = bx^k$ (solved using logarithmic translation $\text{Log } y = k(\text{Log } x) + b$)

Taxa	Comparison	n	k	b	R ²
Ornithomimidae	F, Skull	15	0.773	0.32	0.885
Tyrannosauridae	F, Skull	62	1.028	-0.069	0.948
Ornithomimidae	F, Antor	9	0.938	-0.344	0.899
Tyrannosauridae	F, Antor	36	1.03	-0.274	0.943
Ornithomimidae	F, Orb	8	0.509	0.177	0.809
Tyrannosauridae	F, Orb	38	0.47	0.648	0.559
Ornithomimidae	F, Jaw	11	0.787	0.242	0.862
Tyrannosauridae	F, Jaw	43	0.987	0.038	0.96
Tyrannosauridae	F, Jaw H	55	1.391	-1.873	0.887
Ornithomimidae	F, Hum	14	1.165	-0.614	0.956
Tyrannosauridae	F, Hum	40	0.937	-0.339	0.845
Ornithomimidae	F, Rad	18	1.132	-0.667	0.963
Tyrannosauridae	F, Rad	12	0.828	-0.38	0.907
Ornithomimidae	F, Finger II	15	0.815	0.121	0.729
Tyrannosauridae	F, Finger II	22	0.79	-0.094	0.45
Ornithomimidae	F, Ilium	20	1.074	-0.021	0.899
Tyrannosauridae	F, Ilium	50	1.126	-0.351	0.962
Ornithomimidae	F, P Boot	13	1.055	-0.573	0.928
Tyrannosauridae	F, P Boot	21	1.314	-1.182	0.936
Ornithomimidae	F, Ischium	25	1.157	-0.546	0.929
Tyrannosauridae	F, Ischium	32	1.136	-0.529	0.957
Ornithomimidae	F, Fcirc	34	0.986	-0.499	0.986
Tyrannosauridae	F, Fcirc	72	1.336	-1.413	0.935
Ornithomimidae	F, Tibia	55	1.004	0.025	0.983
Tyrannosauridae	F, Tibia	65	0.782	0.621	0.961
Ornithomimidae	F, Mt III	33	1.013	-0.156	0.943
Tyrannosauridae	F, Mt III	36	0.596	0.972	0.881
Ornithomimidae	F, Toe III	13	1.007	-0.337	0.857
Tyrannosauridae	F, Toe III	21	0.825	0.128	0.789

Deinocoelurus measurements (Supplementary Data) were compared with the curves of these two lineages to determine whether or not the dimensions of this animal were expected as the result of continued growth trajectories of ornithomimosaurids attaining large size. Growth is isometric when $k = 1$, is negatively allometric when $k < 1$, and positively allometric when $k > 1$. Abbreviations: Antor, Antorbital, snout length; b, constant; Skull, length from premaxilla to occipital condyle; F, femur length; Fcirc, minimum femur shaft circumference; Finger II, second manual digit length; Hum, humerus length; Jaw, mandibular length; Jaw H, mandibular height; k, allometric coefficient; Mt III, metatarsal III; n, sample size; Orb, orbit length; P Boot, pubic boot; R², coefficient of determination; Rad, radius; Toe III, sum of phalangeal lengths of pedal digit III.

Comparative population genomics in animals uncovers the determinants of genetic diversity

J. Romiguier^{1,2}, P. Gayral^{1,3}, M. Ballenghien¹, A. Bernard¹, V. Cahais¹, A. Chenuil⁴, Y. Chiari⁵, R. Dernat¹, L. Duret⁶, N. Faivre¹, E. Loire¹, J. M. Lourenco¹, B. Nabholz¹, C. Roux^{1,2}, G. Tsagkogeorga^{1,7}, A. A.-T. Weber⁴, L. A. Weinert^{1,8}, K. Belkhir¹, N. Bierne¹, S. Glémin¹ & N. Galtier¹

Genetic diversity is the amount of variation observed between DNA sequences from distinct individuals of a given species. This pivotal concept of population genetics has implications for species health, domestication, management and conservation. Levels of genetic diversity seem to vary greatly in natural populations and species, but the determinants of this variation, and particularly the relative influences of species biology and ecology versus population history, are still largely mysterious^{1,2}. Here we show that the diversity of a species is predictable, and is determined in the first place by its ecological strategy. We investigated the genome-wide diversity of 76 non-model animal species by sequencing the transcriptome of two to ten individuals in each species. The distribution of genetic diversity between species revealed no detectable influence of geographic range or invasive status but was accurately predicted by key species traits related to parental investment: long-lived or low-fecundity species with brooding ability were genetically less diverse than short-lived or highly fecund ones. Our analysis demonstrates the influence of long-term life-history strategies on species response to short-term environmental perturbations, a result with immediate implications for conservation policies.

Since the early studies of evolutionary genetics, there has been no understanding of how and why genetic diversity levels vary between species. This old puzzle, considered four decades ago as ‘the central problem in population genetics’¹, is still essentially unsolved in the genomic era². Meanwhile, there is increasing evidence that genetic diversity is central to many conservation challenges, such as species response to environmental changes, ecosystem recovery, and the viability of recently endangered populations^{3–7}. In this context, our ability to understand and predict this key aspect of biodiversity seems critical. But is it possible to quantify the contributory ecological and genetic factors? How predictable is the level of genetic diversity of a given species?

Population genetic theory states that neutral genetic polymorphism (that is, diversity) increases with effective population size, N_e , which in a panmictic population is equal to the number of individuals contributing to reproduction. One would therefore expect the genetic diversity of a species to be linked to biological traits associated with abundance, such as body size or fecundity. However, this intuitive prediction has not yet been clearly confirmed by empirical data^{2,8–10}. This is typically explained by invoking the many confounding factors potentially affecting genetic polymorphism, such as mutation rate, population structure, population bottlenecks, selective sweeps, and, more generally, ecological disturbances^{11,12}. Whether demographic or adaptive, historical contingency is often considered to be the main driver of genetic diversity¹¹. According to this viewpoint, polymorphism levels would be expected to fluctuate in time more or less randomly, irrespective of life-history traits.

In the absence of compelling empirical evidence, the relative importance of species biology and ecology (on the one hand) and historical, contingent factors (on the other) in shaping the genetic diversity of species is

still highly contentious. Indeed, current knowledge on species genetic diversity is based on just a handful of model organisms, or small sets of molecular data^{2,8,13}. Various animal taxa and lifestyles, particularly across the invertebrates, have yet to be explored. Here we fill this gap and present the first distribution of genome-wide polymorphism levels across the metazoan tree of life.

We focused on 31 families of animals spread across eight major animal phyla. In each family we produced high-coverage transcriptomic data (RNAseq) for about ten individuals of a particular species. In 25 of these families, we sampled one to three additional species of similar biology and ecology (two to seven individuals each), thus producing taxonomic replicates. The total data set consisted of 374 individual transcriptomes from 76 non-model species (Fig. 1, Extended Data Fig. 1 and Supplementary Tables 1 and 2), from which we predicted protein coding sequences¹⁴ and identified diploid genotypes and single nucleotide polymorphisms^{15,16} (Methods). Across species the number of analysed loci varied from 804 to 20,222 (median 5,347) and the number of polymorphic sites from 1,759 to ~230,000 (median 17,924).

Estimates of the synonymous nucleotide diversity (π_s) spanned two orders of magnitude across species, a range far wider than is usually observed in surveys restricted to a single taxonomic group. The extreme values of π_s were observed in two invertebrate species: 0.1% in the subterranean termite *Reticulitermes grassei*; 8.3% in the slipper shell *Bostrycapulus aculeatus*. Figure 1 illustrates the patchy distribution of low-diversity (green) and high-diversity (red) species across the metazoan phylogeny. It also shows that species π_s values tend to be similar within families, but distinct between families (analysis of variance; $P < 10^{-12}$). Such a strong taxonomic effect would be unexpected if stochastic disturbances and contingent effects were the main drivers of genetic diversity, because species from a given family are not particularly expected to share a common demographic history. Testing this hypothesis more thoroughly, we detected no strong relationship between π_s and any variable related to geography, such as the average distance between GPS records (regression test, $P = 0.19$, $r^2 = 0.02$), maximum distance between GPS records ($P = 0.02$, $r^2 = 0.07$), average distance to Equator ($P = 0.87$, $r^2 = 0.0003$), population structure (measured as F_{it} , $P = 0.22$, $r^2 = 0.02$), invasive status (Student’s t -test, $P = 0.14$) and marine versus continental environment (Student’s t -test, $P = 0.52$).

To test whether species biology can explain variations in π_s , we collected data for six life-history traits potentially related to N_e : adult size, body mass, maximum longevity, adult dispersion ability, fecundity and propagule size (Supplementary Table 3). In contrast to the geographic variables, all these traits were significantly correlated with the nucleotide diversity (Extended Data Fig. 2) and collectively explained 73% of the variance in π_s in a multiple linear regression test ($P < 10^{-10}$). Propagule size, here defined as the size of the stage that leaves its parents and disperses (egg or juvenile depending on species), is by far the most predictive

¹UMR 5554, Institute of Evolutionary Sciences, University Montpellier 2, Centre national de la recherche scientifique, Place E. Bataillon, 34095 Montpellier, France. ²Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland. ³UMR 7261, Institut de Recherches sur la Biologie de l’Insecte, Centre national de la recherche scientifique, Université François-Rabelais, 37200 Tours, France. ⁴Aix-Marseille Université, Institut Méditerranéen de Biodiversité et d’Écologie marine et continentale (IMBE) – CNRS – IRD – UAPV, 13007 Marseille, France. ⁵Department of Biology, University of South Alabama, Mobile, Alabama 36688-0002, USA. ⁶UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, Université Lyon 1, CNRS, 69622 Lyon, France. ⁷The School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK. ⁸Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK.

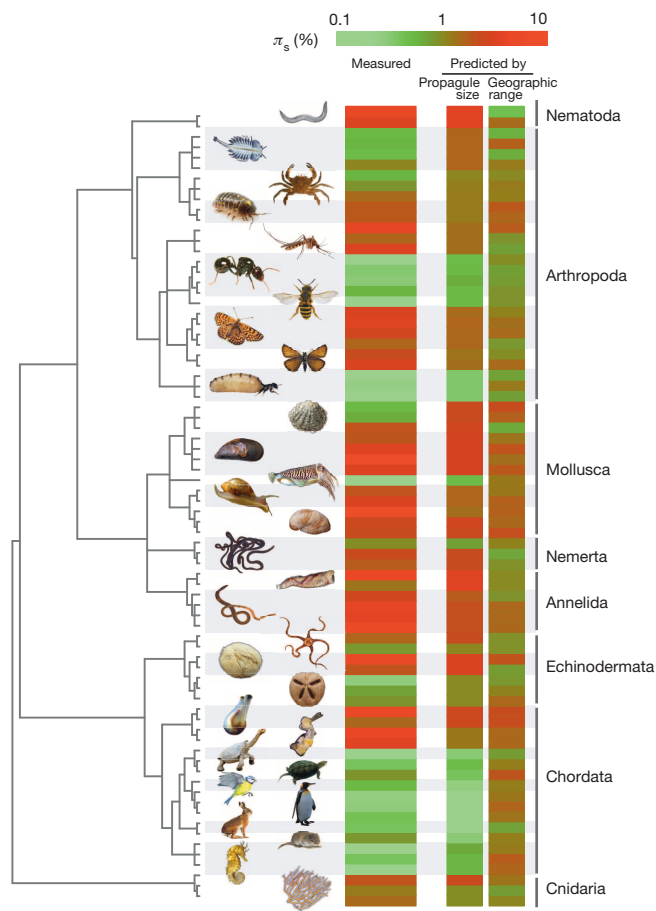


Figure 1 | Genome-wide genetic diversity across the metazoan tree of life. Each branch of the tree represents a species ($n = 76$). The leftmost vertical coloured bar is the estimated genome-wide genetic diversity (π_s), the central bar is the prediction of π_s based on a linear model with propagule size as the explanatory variable ($P < 10^{-14}$, $r^2 = 0.56$), and the rightmost bar is the prediction of π_s based on a linear model with average distance between GPS records, maximal distance between GPS records, average distance to Equator and invasive status as explanatory variables ($P = 0.16$). Each thumbnail corresponds to one metazoan family. Species are in the same order as in Supplementary Table 2 (from top to bottom).

of these variables (linear regression test, $r^2 = 0.56$; Fig. 2a). This is illustrated in Fig. 1 by the good agreement between the observed distribution of π_s (leftmost coloured vertical bar) and the π_s value predicted from propagule size (central bar). The predicted π_s based on four demographic metrics is plotted alongside (rightmost bar) for visual comparison.

We explored in more detail the relative impact on π_s of the various life-history traits of interest here (Extended Data Fig. 2). Figure 2b plots the relationship between π_s and species adult size, a variable typically taken as a proxy for population size in some taxa⁹. Although significant, the correlation is not particularly strong ($P = 0.018$, $r^2 = 0.07$). In particular, species with low genetic diversity cover a large range of body sizes, from less than 1 cm to more than 1 m. Low-polymorphism species include amniotes (turtles, mammals and birds), but also brooding marine species (seahorses, brooding urchins, nemerteans and brittle-stars), eusocial insects (ants, bees and termites) and cuttlefish. These phylogenetically unrelated species have in common a large parental investment in their offspring, as represented in Fig. 2b by the ratio of propagule size to adult size (red). In contrast, species with minimal parental investment (blue) tend to carry high genetic diversity given their size. This is typically the case of highly fecund, broadcast spawning sessile species (such as mussels, non-brooding urchins, nemerteans and brittle-stars, sea squirts and gorgonians). The trade-off between offspring quantity (fecundity) and

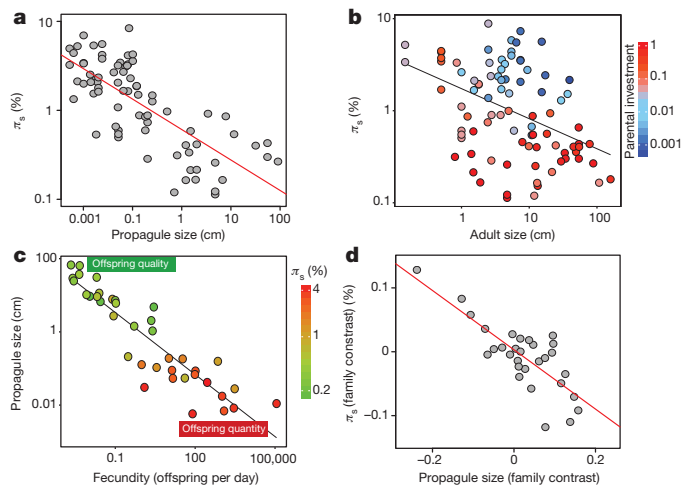


Figure 2 | Life-history traits and genetic diversity relationships. **a**, Relationship between propagule size and π_s ($P < 10^{-14}$, $r^2 = 0.56$, 76 species included; see Fig. 1). **b**, Relationship between adult size and π_s ($P < 0.05$, $r^2 = 0.07$, 76 species included). The colour scale represents the degree of parental investment, here defined as the ratio of propagule size to adult size. **c**, Effect of fecundity per day (x axis) and propagule size (y axis) on genetic diversity (colour scale; $P < 10^{-6}$, $r^2 = 0.69$, 29 family-averaged data points). **d**, Phylogenetic contrasts of family-averaged π_s versus family-averaged propagule size ($P < 10^{-6}$, $r^2 = 0.62$).

quality (propagule size) seems to be the most relevant factor explaining variations in polymorphism between species in the animal kingdom (Fig. 2c). We shall for simplicity hereafter categorize as *K*-strategists the species that tend to invest in the quality of their progeny, and as *r*-strategists those that favour quantity¹⁷.

The correlation we report between life-history traits and π_s is not due to phylogenetic non-independence of the sampled species: taking family averages from Fig. 1 increased the correlation coefficients (from $r^2 = 0.56$ to $r^2 = 0.66$ with propagule size alone, from $r^2 = 0.73$ to $r^2 = 0.79$ with the six life-history traits). When we took into account the between-family phylogenetic tree using independent contrasts, this still resulted in highly significant correlations between π_s and life-history traits ($r^2 = 0.62$ for propagule size; Fig. 2d and Extended Data Fig. 3). These relationships were also unaffected by sampling strategy, sequencing depth, gene expression levels or contaminants (Methods, Supplementary Table 4 and Extended Data Figs 4–6). Finally, our conclusions were unchanged when we included 14 previously published species of mammals¹⁰ or when we restricted the analysis to a subset of common orthologous genes (Supplementary Table 4).

The relationship between π_s and life-history traits, however strong, could in principle be mediated by causative variables that were not included in the analysis. One of these potential confounding factors is the mutation rate: a higher average per-generation mutation rate in *r*-strategists than in *K*-strategists could explain our results irrespective of the population size effect. However, theoretical models and empirical measurements actually support the opposite; that is, an increased per-generation mutation rate in large, long-lived organisms due to a larger number of germline cell divisions per generation and a reduced efficacy of natural selection on the fidelity of polymerases¹⁸. Therefore, as far as we can tell, across-species variations of mutation rate are likely to oppose, not strengthen, the main effect we are reporting here.

We computed the non-synonymous nucleotide diversity, π_n , and this was also found to be correlated with species life-history traits (Extended Data Fig. 2). We found substantial variation in π_n/π_s across metazoan species, and significant correlations with life-history traits, the best predictor in this case being longevity (Extended Data Fig. 7). This positive correlation is predicted by the nearly neutral theory of molecular evolution¹⁹: in small populations (long-lived species), the enhanced genetic drift counteracts purifying selection and promotes the segregation of weakly

deleterious, non-synonymous mutations at high allele frequency. These results also confirm that the relationships we uncovered between life-history traits and diversity patterns are mediated in the first place by an effect of N_e , not of the mutation rate; synonymous and non-synonymous positions being physically interspersed, the π_n/π_s ratio is unaffected by the mutation rate.

Our analysis reveals that polymorphism levels are well predicted by species biology, whereas historical and contingent factors are only minor determinants of the genetic diversity of a species. This unexpected result opens new questions. How can life-history traits be so predictive of π_s in spite of the overwhelming evidence for the impact of ecological perturbations on patterns of genetic variation^{11,12}? Why does the ' r/K gradient' affect genetic polymorphism so strongly?

In an attempt to resolve these paradoxes, we suggest that life-history strategies might influence the response of species to environmental perturbations. Because K -strategy species have been selected for survival and the optimization of offspring quality in complex, stable environments¹⁷, we speculate that they might experience fewer occasional disturbances (or be less sensitive to them), thus ensuring the long-term viability of even small populations. In contrast, only species with a large population-carrying capacity could sustain the 'riskier' r -strategy in the long term, thus buffering the frequent bottlenecks experienced in the context of high environmental sensitivity (see Supplementary Equations for a model formalizing these arguments). According to this hypothesis, environmental perturbations would be a common factor affecting every species, but their demographic impact would depend on the life-history strategy of each species.

This study highlights the importance of species life-history strategy when it comes to turning genetic diversity measures into conservation policy. So far, conservation efforts have mainly been focused on large-sized vertebrates. Here we show that these popular animals represent only a subset of the existing low-diversity, K -strategists. Invertebrate species with strong parental investment are probably equally vulnerable to genetic risks. Our results also indicate that r -strategists will typically show elevated amounts of genetic diversity irrespective of their current demography, which suggests that species of this kind might face significant extinction risks²⁰ without any warning genetic signal.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 March; accepted 17 July 2014.

Published online 20 August; corrected online 12 November 2014 (see full-text HTML version for details).

1. Lewontin, R. *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, 1974).
2. Leffler, E. M. *et al.* Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10**, e1001388 (2012).
3. Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 19–23 (2002).
4. Reusch, T. B. H., Ehlers, A., Hämmerli, A. & Worm, B. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proc. Natl Acad. Sci. USA* **102**, 2826–2831 (2005).
5. Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N. & Vellend, M. Ecological consequences of genetic diversity. *Ecol. Lett.* **11**, 609–623 (2008).
6. Johnson, W. E. *et al.* Genetic restoration of the Florida panther. *Science* **329**, 1641–1645 (2010).
7. Nair, P. Conservation genomics. *Proc. Natl Acad. Sci. USA* **111**, 569 (2014).

8. Bazin, E., Glémin, S. & Galtier, N. Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570–572 (2006).
9. Nabholz, B., Mauffrey, J.-F., Bazin, E., Galtier, N. & Glémin, S. Determination of mitochondrial genetic diversity in mammals. *Genetics* **178**, 351–361 (2008).
10. Perry, G. *et al.* Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res.* **22**, 602–610 (2012).
11. Banks, S. C. *et al.* How does ecological disturbance influence genetic diversity? *Trends Ecol. Evol.* **28**, 670–679 (2013).
12. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
13. Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* **29**, 51–63 (2014).
14. Cahais, V. *et al.* Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol. Ecol. Resources* **12**, 834–845 (2012).
15. Tsagkogeorga, G., Cahais, V. & Galtier, N. The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. *Genome Biol. Evol.* **4**, 740–749 (2012).
16. Gayral, P. *et al.* Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet.* **9**, e1003457 (2013).
17. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
18. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010).
19. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
20. Pinsky, M. L. *et al.* Unexpected patterns of fisheries collapse in the world's oceans. *Proc. Natl Acad. Sci. USA* **108**, 8317–8322 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the following for providing samples: F. Delsuc, E. Douzery, M. Tilak, G. Dugas, S. Harispe, C. Benoist, D. Bouchon (woodlice), J. Bierne, M. Bierne, B. Houseaux, M. Strand, C. Lemaire, D. Lallias, Service Modèle Biologique Station Marine Roscoff (nemertines), X. Turon, S. Lopez-Legentil (Cystodytes), P. Jarne, P. David, R. Dillon, J. Auld, R. Relyea, C. Lively, J. Jokela, V. Poullain, T. Stewart (snails), S. Lapègue, V. Boulo, F. Batista, D. Lallias, L. Fast Jensen, M. Cantou (oysters), J. Do Nascimento, C. Daguin-Thiébaud, M. Cantou (crabs), L. Bonnaud (cuttlefish), D. Aurelle (gorgonians), F. Viard, Y. Pechenik, A. Cahill, R. Collins (slipper limpets), L. Dupont (earthworms), D. Jollivet (trumpet worms), M. A. Felix, I. Nuez (nematodes), N. Rodes, T. Lenormand, E. Flaven (brine shrimps), Rotterdam Zoo, Zurich Zoo, C. Libert, Montpellier Zoo, S. Martin, la Ferme aux Crocodiles, O. Verneau, C. Ayres, M. Carretero, M. Vanberger, K. Pobolsaj, M. Zuffi, C. Palacios, L. du Preez, B. Halpern, Budapest Zoo (turtles), P. Peret, C. Doutrelant, B. Halpern, B. Rosivall (tits), M. de Dinechin, B. Rey (penguins), Z. Melo-Ferreira, P. Alves (hares), N. Brand, M. Chapuisat (bees), R. Blatrix, A. Lenoir, I. Nodet, A. Lugagne, S. Blanquart, L. Serres-Giardi, V. Roustang, N. François, G. Ballantyne, A. Carbonnel, Y. Samuel, G. James, G. Kalytta, F. Guerrini, S. Stenzel, J. Beekman, X. Cerda, S. Ikonen (ants), I. Hanski, S. Ikonen, J. Kullberg, Z. Kolev (fritillary butterflies), F. Viard, X. Turon, Di Jiang, D. Chourrout, B. Vercaemer, E. Newman-Smith, Ascidian Stock Center, Service Modèle Biologique Station Marine Roscoff (ciona), L. Excoffier, G. Heckel (voles), F. Dedeine (termites), C. Atyame, O. Duron, M. Weill (mosquitoes), M. Cantou, H. Violette, F. Batista, J. Hondeville (seahorses), C. Fraisse, G. Pogson, N. Saarnan, J. Normand (mussels), E. Poulin, C. Gonzalez-Weivar, and J. P. Feral (sea urchins). This work was supported by European Research Council advanced grant 232971 (PopPhyl).

Author Contributions N.G. conceived the project. P.G., M.B., N.F., Y.C., L.A.W., G.T., A.C., A.W., J.R., N.G. and N.B. performed sampling and laboratory work. A.B., V.C., E.L., J.R., J.M.L., C.R., P.G., G.T., B.N., R.D., K.B., S.G. and N.G. developed the data analysis pipeline. J.R. collected life-history/geographic variables and produced figures. J.R., A.B., V.C., L.D., E.L. and N.G. analysed the data. S.G., N.B., B.N., J.R. and N.G. provided interpretations and models. J.R., N.B., S.G. and N.G. wrote the paper.

Author Information Data sets are freely available from the Sequence Read Archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>) under project ID SRP042651 and from the Datasets section of the PopPhyl website (<http://kimura.univ-montp2.fr/PopPhyl/>), in which predicted single nucleotide polymorphisms and genotypes are provided as .vcf files. Scripts and executable files are freely available from the Tools section of the PopPhyl website. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.G. (nicolas.galtier@univ-montp2.fr).

METHODS

Sampling and sequencing. The 76 analysed species were selected based on phylogenetic and ecological criteria with the goal of gathering a panel representative of the metazoan diversity. In each species, from two to eleven individuals were collected in various localities of their natural geographic range (Supplementary Table 1) or from zoos (*Chelonoidis nigra*). Depending on the species, the whole body, body parts, organs or tissues were dissected and preserved in RNA later, flash-frozen, or processed immediately (Supplementary Table 2). For each sample, total RNA was extracted using standard and improved protocols²¹, and a non-normalized complementary DNA library was prepared. The libraries were sequenced on a Genome Analyzer II or HiSeq 2000 (Illumina) to produce 100-base-pair (bp) or 50-bp single-end fragments. In 12 species, an additional normalized random-primed cDNA library was prepared and sequenced for half a run using a 454 Genome Sequencer (GS) FLX Titanium Instrument (Roche Diagnostics). Illumina reads from two to four individuals in 14 mammalian species¹⁰ were downloaded from the SRA database. Reads were trimmed of low-quality terminal portions with the SeqClean program (<http://compbio.dfci.harvard.edu/tgi/>). The fastQ program was applied to Illumina reads and revealed only a limited amount of motif enrichment: the number of motifs in significant excess varied between 0 and 17 across species, its median being 1.

Transcriptome assembly, read mapping, coding sequence prediction. *De novo* transcriptome assembly based on the 454 (when available) and Illumina reads was performed by following strategies B and D in ref. 14, using a combination of the programs Abyss and Cap3. Illumina reads were mapped to predicted cDNAs (contigs) with the BWA program. Contigs with a per-individual average coverage below $\times 2.5$ were discarded. Open reading frames (ORFs) were predicted with the Trinity package. Contigs carrying no ORF longer than 200 bp were discarded. In contigs including ORFs longer than 200 bp, 5' and 3' flanking non-coding sequences were deleted, thus producing predicted coding sequences that are hereafter referred to as loci.

Calling single nucleotide polymorphisms (SNPs) and genotypes. At each position of each locus and for each individual, diploid genotypes were called according to the method described in ref. 15 (model M1) and improved in ref. 16, using the reads2snps program. This method first estimates the sequencing error rate in the maximum-likelihood framework, calculates the posterior probability of each possible genotype, and retains genotypes supported at $>95\%$; otherwise missing data are called. A minimum of ten reads per position and per individual were required to call a genotype. Then polymorphic positions were filtered for possible hidden paralogs (duplicated genes), using a likelihood ratio test based on explicit modelling of paralogy¹⁶. The across-species average percentage of SNPs discarded for suspicion of paralogy was 7.65%. Positions at which a genotype could not be called in a sufficient number of individuals were discarded. Calling k the number of sampled individuals for a given species ($2 \leq k \leq 11$), the minimum number of genotyped individuals required to retain a position was set to $k/2$ when $k > 5$, to $k - 1$ when $k = 4$ or $k = 5$, and to k when $k < 4$.

Control for contamination. Each locus of each species was translated to protein and compared with the non-redundant NR database using BlastP in search for possible contaminants. The percentage of loci for which no significant hit (e-value < 0.001) was retrieved varied greatly between species, reflecting the taxonomic representation of sequences in NR. The percentage of no-hits was below 10% in mammals, but reached values above 20% in echinoderms and cnidarians. When one or several hits were found, the GenBank taxonomy of the first hit was recorded. Overall, 98.7% of first hits were assigned to Metazoa. The percentage of non-metazoan first hits was below 2% in 63 species out of 76, reaching its maximum (5.5%) in the trumpet worm *Pectinaria koreni*. Contamination by known microbes therefore seems negligible in our data set. Extended Data Figure 6 displays the taxonomic distribution of hits for four representative species—a mammal, an insect, a mollusc and an annelid. The results of our analyses were qualitatively unchanged when we removed loci that hit a non-metazoan and/or no-hit loci. In all cases, linear regression tests between π_s and propagule size yielded the same r^2 of 0.55 ($P < 10^{-13}$) as in our main analysis.

Life-history, ecological and geographical variables. Species life-history traits (adult size, juvenile size, body mass, longevity, fecundity and adult speed) were retrieved from the literature (Supplementary Table 3). Invasive/non-invasive status was obtained from the Global Invasive Status Database (<http://www.issg.org/database/species/List.asp>). The Global Biodiversity Information Facility database (<http://www.gbif.org/>) was queried to retrieve the GPS records corresponding to documented observations of individuals from the species of interest here. These data were merged with the GPS records of our own samples. For each species, the average and maximum distance between two distinct GPS records and the average distance to the Equator were computed (Supplementary Table 2).

Statistical analyses. Population genomic statistics π_s , π_n and F_{it} were calculated by using home-made programs that rely on the Bio++ libraries²². The F_{it} calculation was corrected for small sample sizes in accordance with ref. 23. Confidence intervals were obtained by bootstrapping loci. Regression analyses were conducted in R. Variables were log-transformed before linear regressions were performed. The linear model

including π_s and propagule size alone met the required assumptions of normally distributed residual errors (Shapiro's test, $P = 0.19$) and homoscedasticity (Fligner–Killeen's test, $P = 0.48$). The same remark is valid for the multiple linear model including π_s and the six life-history traits (Shapiro's test, $P = 0.31$; Fligner–Killeen's test, $P = 0.49$). Family-level phylogenetic independent contrast analysis was performed with the APE package based on the tree shown in Extended Data Fig. 3, in which branch lengths are proportional to time. Divergence time estimates were retrieved from the TimeTree database (expert result, or average value if expert result was missing). When divergence time estimates were not available (Polycitoridae–Cionidae, Hesperidiidae–Nymphalidae, Calyptraeidae–Physidae, Mytilidae–Ostreidae), they were inferred on the basis of the divergence dates of neighbouring nodes.

SNP calling quality controls. The main analyses of this study were reproduced in three ways: first, with an increased minimum number of reads per position per individual of 30 instead of 10, second, removing five bases from each end of each read, and third, not using 454 data, thus controlling for a potential effect of insufficient sequencing depth, low-quality base calls near read ends and sequencing technology. In all three cases the results were highly similar to the main analysis (Supplementary Table 4; columns 'depth = 30X', 'clip_ends' and 'no_454', respectively), indicating that the analysis was robust to these technical caveats. No difference in π_s was detected between species showing versus not showing a significant excess of certain motifs by fastQC.

GC content. In each species, the correlation coefficient between contig GC content and contig π_s was calculated. It was significantly positive in 37 species, significantly negative in 18 species, and non-significantly different from zero in 21 species. The squared correlation coefficients (r^2) were relatively low (median r^2 0.007; maximum r^2 0.16 in *Physa acuta*), suggesting only a weak effect of GC content on π_s . For each species, the average contig GC content was calculated and correlated to the average π_s or propagule size. No significant relationship was detected, which indicates that the variation in GC content across genes and across species has no substantial impact on the results of this study.

Individual and locus sampling. No significant relationship was found between π_s and the number of sampled individuals per species, or between π_s and the number of sampled loci per species (Extended Data Fig. 4). The robustness to sampling of the relationship between propagule size and π_s was further assessed in two ways. First, for each species, loci were randomly subsampled. Extended Data Figure 5 displays the squared coefficient of correlation between propagule size and π_s as a function of the per species number of analysed loci. It shows that as few as 50 loci are enough to capture the relationship with a good probability. Second, for each species, only two individuals were randomly selected and the analyses were conducted again. Results were highly similar to the main analysis: the relationship between propagule size and π_s was unchanged and highly significant ($P < 10^{-15}$, $r^2 = 0.55$), thus indicating that population sample size is not an issue.

Orthologous genes. The coding sequences of 129 genes or gene fragments previously identified as orthologous across metazoans²⁴ (hereafter called 'core genes') were downloaded. In each of our species, contigs predicted to be orthologous to one of the core genes by reciprocal best BLAST hit were selected (expected e-value 0.0001, hits with a number of identical matches less than 80 and a bitscore of less than 1,200 were discarded). The number of such predicted core gene orthologues varied between 40 and 122 among species. We restricted the data set to the 39 species including at least 21 core gene orthologues, and reproduced the analysis. Colon 'orthologues' from Supplementary Table 4 shows that the correlation between π_s and life-history traits was still strong and significant when a subset of common genes was considered.

Expression level. In each species, the expression level of each locus was estimated as the average number of bases read per position. Correlating π_n/π_s to expression level across genes revealed no significant relationship in 33 species, and weak relationships ($r^2 < 0.27$) in 57 species. The relationship between π_n/π_s and expression level, when detected, was negative, as expected under the hypothesis of a stronger selective pressure acting on high-expressed genes. Then, for each species, loci were grouped into three equal-sized bins of genes with high, medium or low expression. Each of these categories, taken separately, provided a strong correlation between species propagule size and π_s ($r^2 = 0.57$, 0.62 and 0.62, respectively).

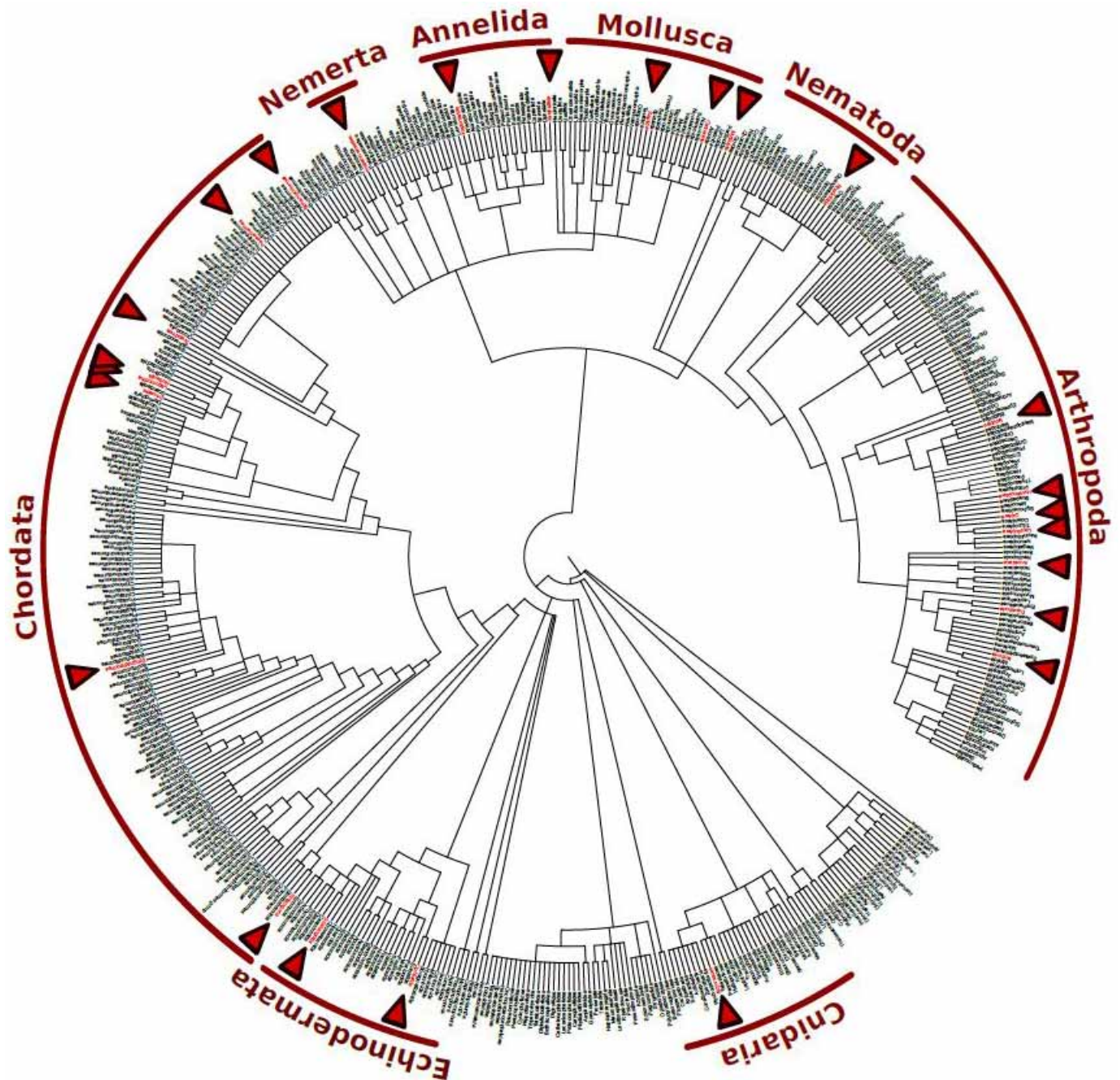
Linkage. The diversity of a neutral locus might be affected by selection at linked sites. This is particularly true of synonymous sites, which reside within coding sequences; that is, targets for natural selection. One would therefore predict a lower genetic diversity in species experiencing a low genomic average recombination rate. We lacked a recombination map in most of the analysed species; we therefore relied on generic taxonomic patterns to approach this issue. Eusocial hymenopterans are known to experience a recombination rate one order of magnitude higher than most animals²⁵. In contrast, dipteran Culicidae (mosquitoes) experience relatively small amounts of recombination (median recombination rate in eusocial hymenoptera, 9.7 centimorgans per megabase (cM Mb⁻¹); median recombination rate in Culicidae, 0.3 cM Mb⁻¹)²⁵. The linkage effect would therefore predict a decreased π_s in Culicidae and an elevated π_s in eusocial hymenopterans. We observed the opposite: π_s varied from 0.0016 to 0.0058 in our five eusocial hymenopteran species, which is below the average metazoan

π_s (0.015), and one order of magnitude below the π_s of our three Culicidae species (0.016–0.041). This result, which is consistent with the propagule-size hypothesis, does not suggest that the between-species variation in genomic average recombination rate strongly influences our results.

Population structure. The genetic distance between individuals was defined as $(H_b - H_w)/H_w$, where H_b is the probability of drawing two distinct alleles when sampling one copy from each of the two considered individuals, and H_w is the average heterozygosity of the two considered individuals. In species containing more than four individuals, the genetic distance was calculated for each pair of individuals and correlated to the geographic distance; the squared coefficient correlation, r^2 , which measures genetic isolation by distance, ranged from 0.0008 to 0.73 (Supplementary Table 5). Consistent with the phylogeographic literature, it was high ($r^2 > 0.35$) and significant in, for example, *Ciona intestinalis* A, *Melitaea cinxia* and *Sepia officinalis*, and low ($r^2 < 0.02$, n.s.) in, for example, *Culex pipiens*, *Lepus granatensis* and *Crepidula fornicata*. The F_{it} statistic was significantly higher, on average, in terrestrial (median $F_{it} = 0.25$) than in marine (median $F_{it} = 0.02$) species (t -test, $P = 0.029$) when only species including at least five individuals were considered. No significant relationship was detected between π_s and absolute values of F_{it} ($P = 0.22$, $r^2 = 0.05$, only species with more than four individuals included), which does not suggest any confounding effect of population structure in our analysis.

Ethical statement. Living animals were manipulated according to the 'Charte Nationale Portant sur l'Éthique de l'Expérimentation Animale'. Sampling of protected species was performed under permits 53/2009 (Galicia, Spain, *Emys orbicularis*), 2009/11/12 (Aude, France, *Emys orbicularis*), 503/05/07/2006 (Pisa, Italy, *Emys orbicularis*), 35601-60/2005-4 (Slovenia, *Emys orbicularis*), and 009-01-1230/a34-455 (France, *Parus caeruleus*). *Aptenodytes patagonicus*, *Eudyptes moseleyi* and *Eudyptes filholi* individuals were sampled by Institut Polaire Français Paul Emile Victor, program IPEV 131. *Chelonoidis nigra* individuals were handled and sampled by the veterinarians and staff of the Zurich zoo (Switzerland), Rotterdam zoo (the Netherlands), and A Cupulatta zoo (France) in accordance with the Code of Practice and Code of Ethics established by the European Association of Zoos and Aquaria.

21. Gayral, P. *et al.* Next-generation sequencing of transcriptomes: a guide to RNA isolation in non-model animals. *Mol. Ecol. Resources* **11**, 650–661 (2011).
22. Guéguen, L. *et al.* Bio++: efficient, extensible libraries and tools for molecular evolution. *Mol. Biol. Evol.* **30**, 1745–1750 (2013).
23. Weir, B. S. & Cockerham, C. C. Estimating F -statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
24. Philippe, H. *et al.* Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009).
25. Wilfert, L., Gadau, J. & Schmid-Hempel, P. Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity* **98**, 189–197 (2007).

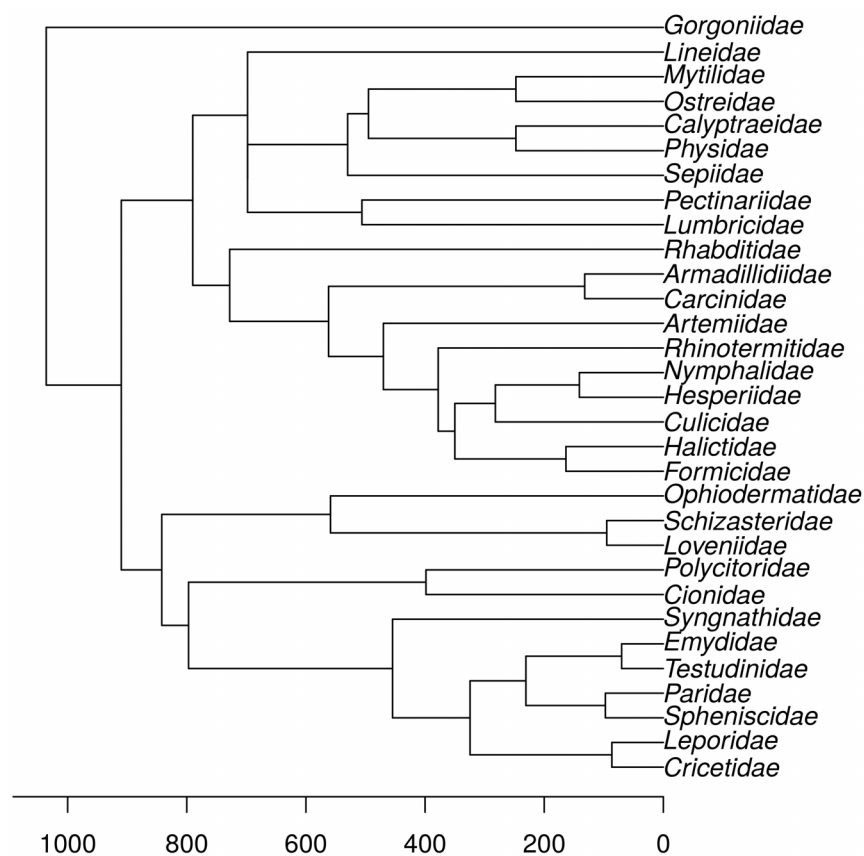


Extended Data Figure 1 | Phylogenetic tree of metazoan orders and the position of the taxa analysed in this study. The tree topology is consistent with the NCBI taxonomy. Red arrows identify 25 orders that were sampled. Five gastropod species from two distinct families (Calyptraeidae (*Crepidula*

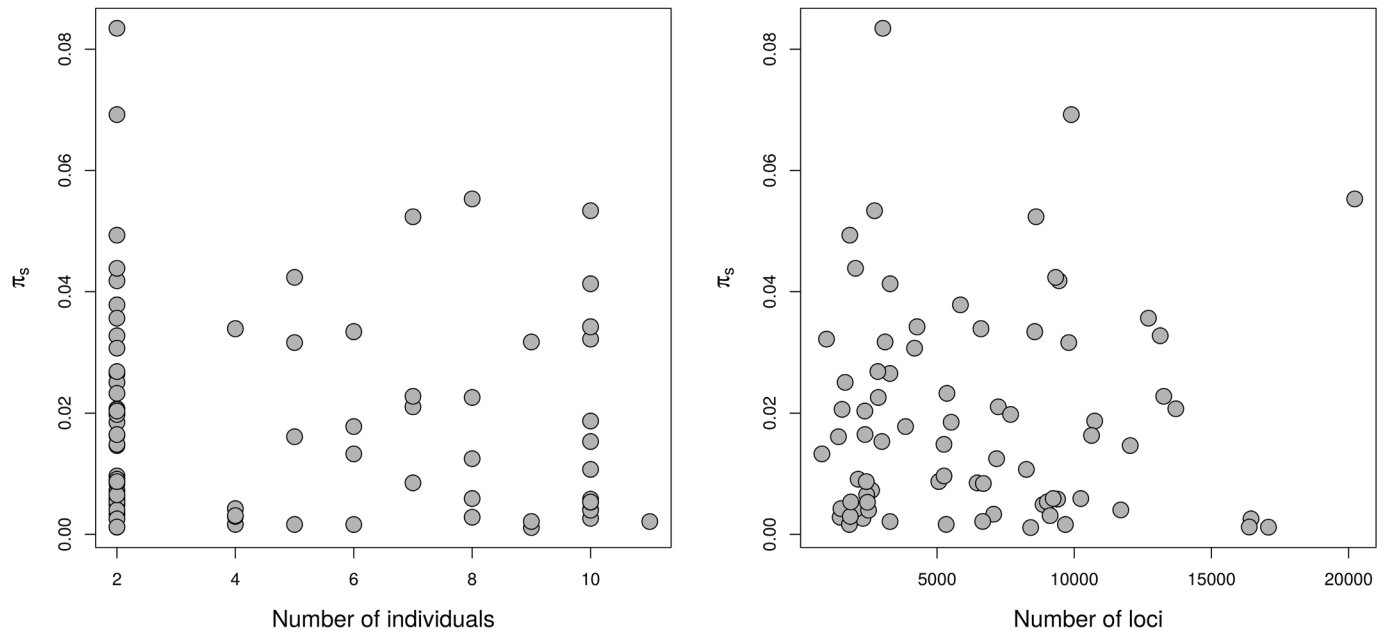
forficata, *C. plana* and *Bostrycapulus aculeatus*) and Physidae (*Physa acuta* and *P. gyrina*)) are not represented because they lacked any assignment to an order in current taxonomy.



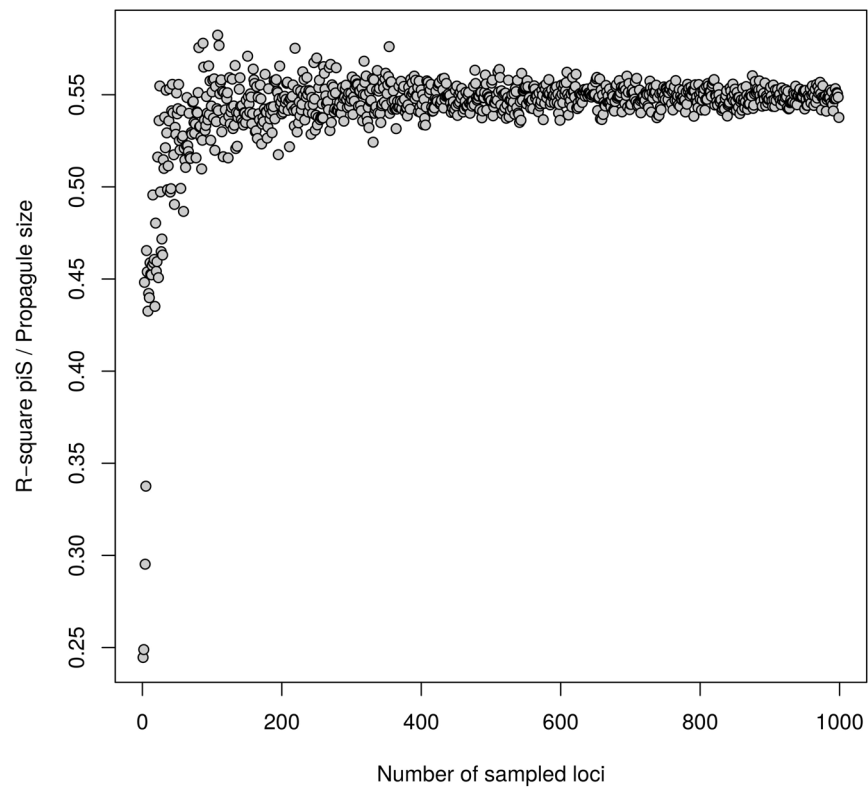
Extended Data Figure 2 | Correlations between genetic diversity and life history variables. Blue indicates a positive relationship, red a negative one; colour intensity is proportional to Pearson's correlation coefficient.



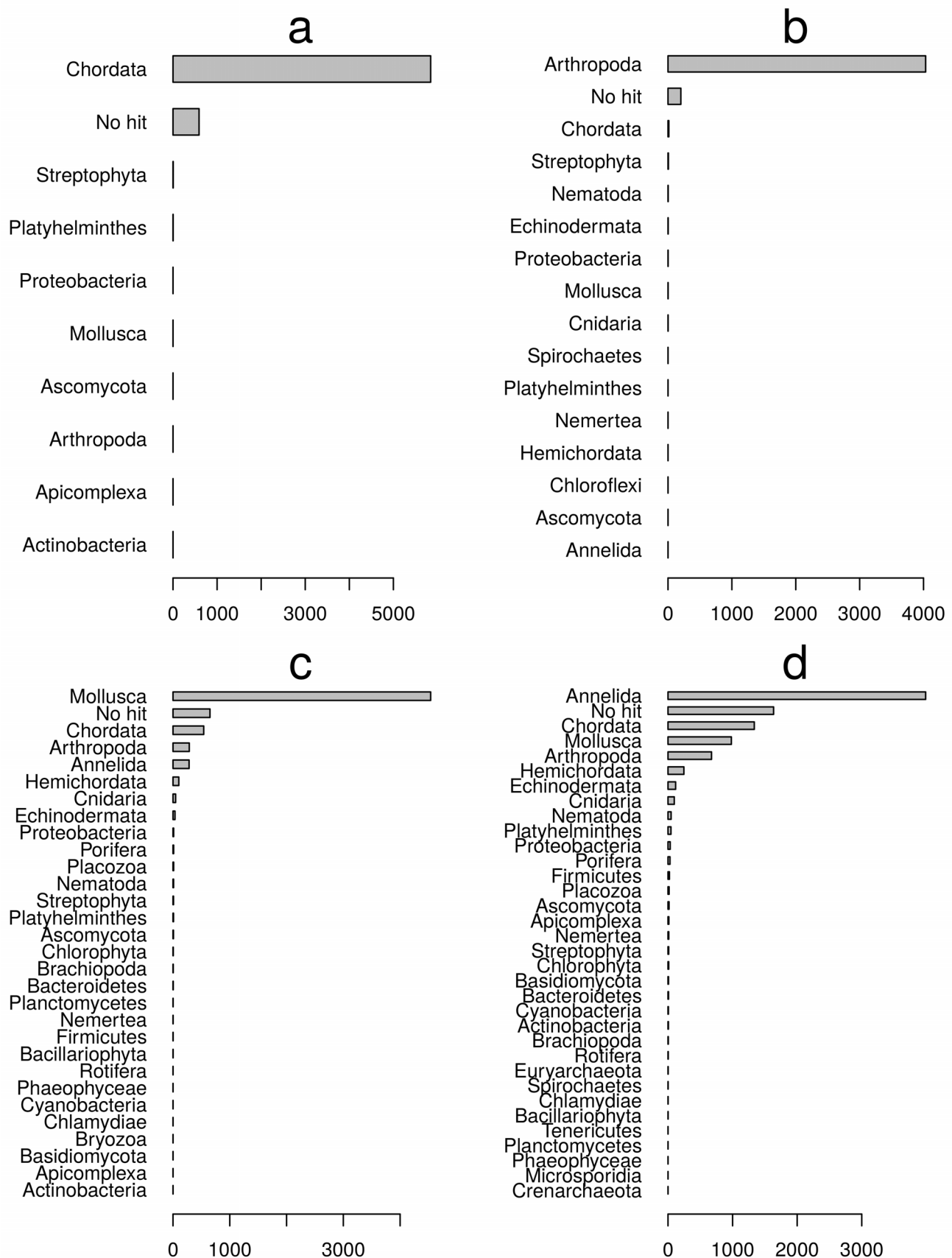
Extended Data Figure 3 | Family-level phylogenetic tree (31 families included). The scale is in million years of divergence.



Extended Data Figure 4 | Absence of significant correlation between species genetic diversity with individual sampling size ($P = 0.47$, $r^2 = 0.007$) and locus sampling size ($P = 0.78$, $r^2 = 0.001$).

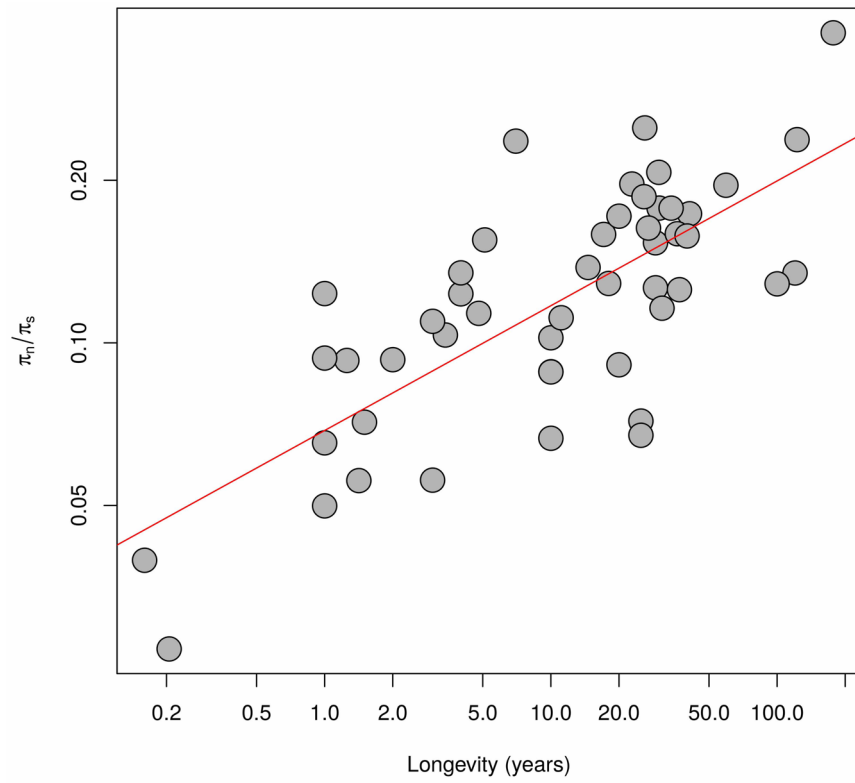


Extended Data Figure 5 | Relationship between the π_s /propagule-size r^2 and the number of sampled loci.



Extended Data Figure 6 | Phylum distribution of the first BLAST hit in four representative species. **a**, Common vole (*Microtus arvalis*). **b**, Glanville

fritillary butterfly (*Melitaea cinxia*). **c**, Blue mussel (*Mytilus edulis*). **d**, Earthworm (*Allolobophora chlorotica*).



Extended Data Figure 7 | Correlation between π_n/π_s and maximum longevity ($P < 10^{-8}$, $r^2 = 0.54$). Only species with at least four individuals are included.

Radial glia require PDGFD–PDGFR β signalling in human but not mouse neocortex

Jan H. Lui^{1*†}, Tomasz J. Nowakowski^{1*}, Alex A. Pollen¹, Ashkan Javaherian^{1†}, Arnold R. Kriegstein^{1§} & Michael C. Oldham^{1§}

Evolutionary expansion of the human neocortex underlies many of our unique mental abilities. This expansion has been attributed to the increased proliferative potential^{1,2} of radial glia (RG; neural stem cells) and their subventricular dispersion from the periventricular niche^{3–5} during neocortical development. Such adaptations may have evolved through gene expression changes in RG. However, whether or how RG gene expression varies between humans and other species is unknown. Here we show that the transcriptional profiles of human and mouse neocortical RG are broadly conserved during neurogenesis, yet diverge for specific signalling pathways. By analysing differential gene co-expression relationships between the species, we demonstrate that the growth factor *PDGFD* is specifically expressed by RG in human, but not mouse, corticogenesis. We also show that the expression domain of *PDGFR β* , the cognate receptor^{6,7} for *PDGFD*, is evolutionarily divergent, with high expression in the germinal region of dorsal human neocortex but not in the mouse. Pharmacological inhibition of *PDGFD*–*PDGFR β* signalling in slice culture prevents normal cell cycle progression of neocortical RG in human, but not mouse. Conversely, injection of recombinant *PDGFD* or ectopic expression of constitutively active *PDGFR β* in developing mouse neocortex increases the proportion of RG and their subventricular dispersion. These findings highlight the requirement of *PDGFD*–*PDGFR β* signalling for human neocortical development and suggest that local production of growth factors by RG supports the expanded germinal region and progenitor heterogeneity of species with large brains.

Radial glia are the physical substrate⁸ and progenitor population that underlie production of most cells in human neocortex². We sought to determine a general transcriptional ‘signature’ of human neocortical RG (hRG) as a starting point for identifying genes that may regulate uniquely human aspects of cortical development. We and others have previously shown that gene co-expression analysis of heterogeneous tissue samples can deconvolve transcriptional signatures of distinct cell types without cell isolation or purification^{9,10}. Because prenatal samples of human neocortex are scarce, we developed a novel strategy called Gene Co-expression Analysis of Serial Sections (GCASS) that exploits variation in cellular abundance across serial sections of a single tissue sample to reveal cell-type-specific patterns of gene expression (Fig. 1a–c and Extended Data Fig. 1; see Supplementary Information for methods, rationale and further discussion). We applied GCASS to 87 150- μ m sections of a single human cortical sample from gestational week 14.5 (GW14.5, corresponding to peak layer V neurogenesis¹¹; Supplementary Table 1) and identified 55 modules of co-expressed genes. Six modules overlapped significantly with a set of genes that we determined were expressed significantly higher in fluorescence-activated cell sorting (FACS)-sorted mouse RG (mRG) versus intermediate progenitor cells (FACS mRG; Extended Data Fig. 1 and Supplementary Table 2), suggesting that they might represent transcriptional signatures of hRG (Fig. 1d). Analysis of laser-microdissected samples from three independent transcriptomic data sets^{12,13} confirmed that genes in these modules are most highly

expressed in the ventricular zone (VZ) and subventricular zone (SVZ) of developing human neocortex, where both ventricular (vRG) and outer subventricular (oRG) subtypes of RG reside⁴ (Extended Data Fig. 2).

To produce a consensus transcriptional signature for GW14.5 hRG, we first summarized each of these six modules by its first principal component/module eigengene^{14,15} (ME) and calculated the Weighted Gene Co-expression Network Analysis¹⁶ measure of intramodular gene connectivity, $k_{ME}^{10,14}$ (see Fig. 1c). k_{ME} quantifies the extent to which a gene

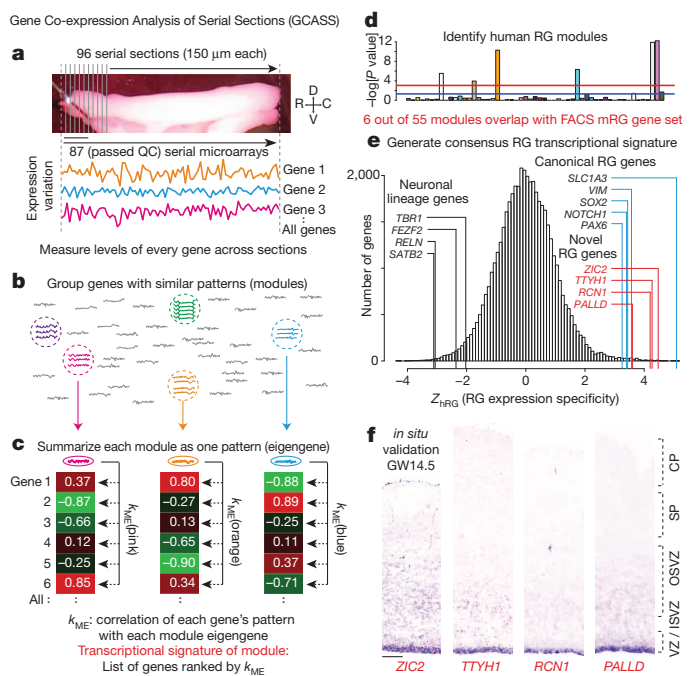


Figure 1 | GCASS identifies a transcriptional signature of radial glia in human neocortex. Conceptual framework is indicated in a–c.

a, Transcriptional profiling of serial sections ($n = 87$, Illumina HT-12 v4 microarrays) from a GW14.5 human neocortical specimen. Scale bar, 2.5 mm. C, caudal; D, dorsal; R, rostral; V, ventral. **b**, Genes with similar expression patterns are grouped into modules, which may reflect cell-type-specific gene co-expression¹⁰. **c**, The transcriptional signature of a module is defined as a list of genes ranked by their correlation to the module eigengene¹⁴. **d–f**, Finding human RG modules. **d**, Six out of fifty-five modules were significantly enriched (Fisher's exact test) with the FACS mRG gene set. Blue line, $P = 0.05$; red line, $P = 9.1 \times 10^{-4}$ (Bonferroni correction). **e**, Genome-wide distribution of predicted GW14.5 neocortical RG expression specificity (Z_{hRG}) based on enriched modules in **d**. **f**, ISH confirms RG expression specificity for novel RG genes from **e**. Scale bar, 300 μ m. CP, cortical plate; ISVZ, inner subventricular zone; OSVZ, outer subventricular zone; SP, subplate. Sample size was determined on the basis of statistical and technical considerations (see Supplementary Information).

¹Department of Neurology and The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, California 94143, USA. [†]Present addresses: Department of Biology and Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA (J.H.L.); Gladstone Institute of Neurological Disease, San Francisco, California 94158, USA (A.J.).

*These authors contributed equally to this work.

§These authors jointly supervised this work.

conforms to the characteristic expression pattern of a module and can predict gene expression specificity for individual cell types¹⁰. k_{ME} values for the six modules were combined into a single measure (Z_{hRG}), with higher values predicting greater expression specificity for hRG (Fig. 1e). Genes with high Z_{hRG} values included markers of neocortical RG such as *SLC1A3* (*GLAST1*), *VIM*, *SOX2*, *NOTCH1* and *PAX6* (Fig. 1e, blue lines). Genes with low Z_{hRG} values included markers of committed neuronal lineages such as *TBR1*, *FEZF2*, *RELN* and *SATB2* (Fig. 1e, black lines). We performed *in situ* hybridization (ISH) and immunostaining on independent prenatal human neocortical samples for genes with high Z_{hRG} values that have not, to our knowledge, previously been implicated in RG biology (Fig. 1e, red lines and Extended Data Fig. 3). In all cases, expression of these genes was restricted to the VZ/SVZ (Fig. 1f and Extended Data Fig. 3). These results indicate that GCASS can discern a general transcriptional signature of hRG from a single, heterogeneous tissue sample without cell labelling, isolation, or purification. Moreover, because the sample derives from a single individual, this strategy implicitly controls for genotype and developmental stage and has broad implications for the molecular analysis of rare tissue samples.

To establish the robustness of the hRG transcriptional signature, we analysed four additional prenatal human cortex gene expression data sets^{12,13} (see also <http://www.brainspan.org/rnaseq/search/index.html>) that were generated with diverse sampling strategies and technology platforms (Extended Data Table 1). In parallel, we analysed three embryonic mouse cortex gene expression data sets^{12,17,18} (largely embryonic day (E)14–14.5, corresponding to peak layer V neurogenesis¹¹; Extended Data Table 1) to establish a robust mRG transcriptional signature. For each data set, we constructed an unsupervised co-expression network, identified the module with the most significant overlap with the FACS mRG gene set, and calculated k_{ME} values ($RG\ k_{ME}$) for every gene with respect to this RG module (Supplementary Table 3). To facilitate comparisons of RG transcriptional signatures, we mapped all probe sets/transcripts to a common identifier (HomoloGene ID) and converted $RG\ k_{ME}$ values for each data set into percentile ranks ($RG\ PR$; workflow schematic, Extended Data Fig. 4). Genome-wide correlations of $RG\ PR$ among all data sets ranged from 0.36 to 0.95 (human) and 0.47 to 0.70 (mouse) ($P < 2.2 \times 10^{-16}$; Extended Data Fig. 5), demonstrating the robustness of RG transcriptional signatures in both species.

To explore the global extent of gene expression conservation between hRG and mRG, we calculated the mean $RG\ PR$ for each gene across the five human ($\overline{hRG\ PR}$) three mouse ($\overline{mRG\ PR}$) and all eight ($\overline{RG\ PR}$) data sets (Supplementary Table 3). For genes present in at least one data set in each species ($n = 15,576$), the correlation between $\overline{hRG\ PR}$ and $\overline{mRG\ PR}$ was 0.48 ($P < 2.2 \times 10^{-16}$; Fig. 2a), indicating broad conservation of transcriptional programs that are active during cortical development in humans and mice. Genes with $\overline{hRG\ PR}$ and $\overline{mRG\ PR} > 80$ (Fig. 2a, +RG box) or < 20 (Fig. 2a, -RG box) were significantly enriched with sets of genes expressed by RG or neurons, respectively (Supplementary Table 4). Furthermore, the most conserved set of core RG genes ($\overline{hRG\ PR}$ and $\overline{mRG\ PR} > 95$; Fig. 2a, green box) included the canonical RG markers *VIM* and *PAX6*, and *SOX* family members implicated in nervous system development such as *SOX2*, *SOX3*, *SOX9* and *SOX21* (Fig. 2b). Also present among this core set were elements of the Notch (*NOTCH2* and *HES1*) and Wnt- β -catenin (*SFRP1*, *FZD8* and *LRP4*) signalling pathways, which critically regulate neurogenesis². In addition, we observed a cohesive subgroup of genes involved in cell cycle regulation, including *CKS2*, *RACGAP1*, *MELK*, *CCNB1*, *CCNA2*, *ASPM* and *MKI67*. Finally, several genes in this group have not, to our knowledge, previously been implicated in RG biology, including *PSAT1*, *DDAH1*, *AIF1L*, *PSRC1* and *ACSS1*. Together, these results provide a framework for identifying conserved and distinct aspects of gene expression in hRG and mRG.

To identify evolutionary changes in RG gene expression that might underlie neurodevelopmental differences between human and mouse, we assessed homologous genes based on predicted differences in RG expression specificity between the species (differential RG specificity

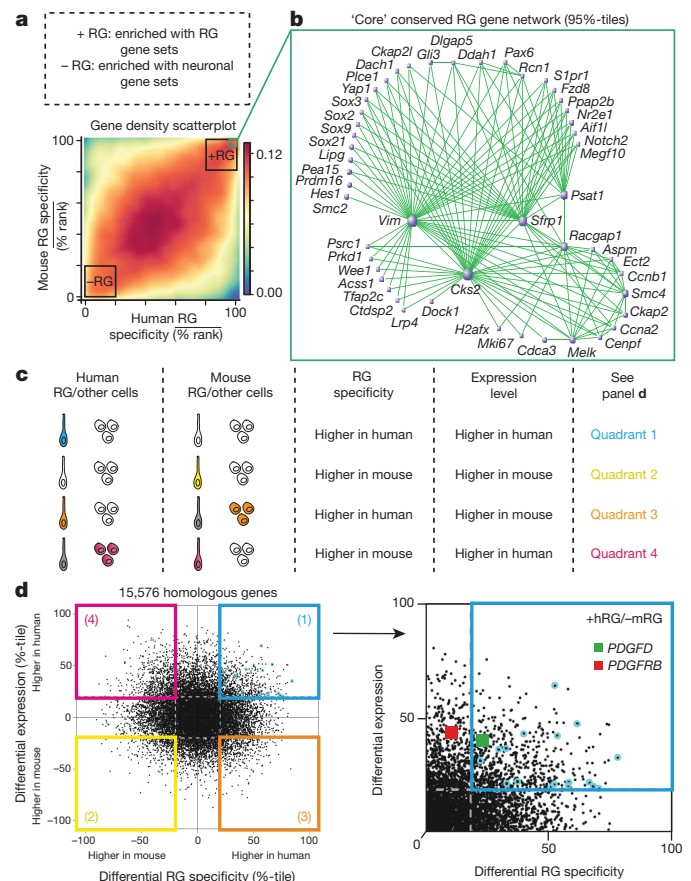


Figure 2 | Combined differential RG specificity/differential expression analysis identifies gene expression differences that distinguish neocortical RG between human and mouse. **a**, Predicted RG expression specificity for 15,576 homologous genes in human and mouse. (See also Supplementary Table 3 and 4.) **b**, The core transcriptional architecture of RG. Node size corresponds to connectivity. **c**, Idealized examples of four types of evolutionary change that could drive differences in predicted hRG and mRG expression specificity. **d**, Comparison of differential RG specificity and differential expression for homologous genes between human and mouse. Dashed grey lines delineate arbitrary thresholds ($-20, 20$) used to define quadrants (1–4) for enrichment analyses (Supplementary Table 4). Cyan circles highlight 18 genes predicted to be expressed in hRG but not mRG; one of these encodes the growth factor *PDGFD* (right; green square). Its cognate receptor *PDGFRB* is also highlighted (right; red square).

(DS) = $\overline{hRG\ PR} - \overline{mRG\ PR}$). Because such differences could emerge from gene expression changes in RG or neighbouring cell types (Fig. 2c), we also compared the relative expression levels of homologous genes between the species (differential expression (DE) = $\overline{hEx\ PR} - \overline{mEx\ PR}$). Genome-wide analysis of DS/DE revealed four distinct quadrants of genes that differed substantially between human and mouse (Fig. 2d). Consistent with the proposed model (Fig. 2c), these quadrants were enriched with sets of genes expressed in distinct patterns in developing human and mouse cortex (Supplementary Table 4).

We reasoned that quadrant 1 might contain genes with 'Boolean' expression differences (that is, ON in hRG and OFF in mRG) required for human but not mouse neocortical development. We applied stringent criteria to identify and validate 18 candidate genes in this quadrant with strong and consistent evidence of expression in hRG and no evidence of expression in mRG: *ABHD3*, *ASAP3*, *BMP7*, *C5*, *C8orf4*, *FAM107A*, *FOXP4*, *ITGA2*, *LRIG3*, *LRRC17*, *PAM*, *PDGFD*, *PDLIM3*, *RFTN2*, *SLC2A10*, *SP110*, *STOX1* and *ZC3H4V1* (Fig. 2d, cyan circles; see also Extended Data Fig. 6). Because secreted growth factors could alter the size of the developing neocortex by influencing proliferation, we focused on *PDGFD* (Fig. 2d (right panel), green square), which has not, to our

knowledge, previously been implicated in cortical development of any species. Compared to genes with the highest RG PR (conserved in both species), *PDGFD* expression was highly correlated in human, but not mouse (Fig. 3a, b). ISH confirmed that *PDGFD* was expressed by RG throughout the VZ of GW14.5 human neocortex (Fig. 3c). In contrast, *Pdgfd* was not detected in RG (or any cell type) in E15.5 mouse neocortex (Fig. 3d). These expression differences were consistent for multiple ages in human (GW14.5–GW18.2) and mouse (E14.5–E17.5) prenatal cortex (Extended Data Fig. 7).

The effects of *PDGFD* are specifically mediated by the *PDGFRβ* receptor, which upon phosphorylation can trigger signalling pathways that promote cell proliferation^{6,7,19}. Although *PDGFRB* did not meet the same stringent criteria as *PDGFD*, its location on the DS/DE plot was proximal to *PDGFD* (Fig. 2d (right panel), red square), also suggesting species differences in *PDGFRB* expression. Compared to genes with the highest RG PR, *PDGFRB* expression was moderately correlated in human, but not mouse (Fig. 3e, f). Immunostaining for *PDGFRβ* in GW14.5 human brain revealed strong expression throughout the telencephalic germinal zones (VZ/SVZ) and in vascular pericytes, with highest

levels in dorsolateral cortical progenitors and the lateral ganglionic eminence (Fig. 3g). In contrast, immunostaining for *PDGFRβ* in E15.5 mouse brain revealed expression in lateral ganglionic eminence progenitors but no evidence of expression in cortical progenitors (Fig. 3h). However, we did observe very low levels of *Pdgfrb* transcript in the VZ of lateral mouse cortex²⁰ (Extended Data Fig. 8), leaving open the possibility of modest, region-specific function. Collectively, these results indicate that expression patterns of *PDGFD* and *PDGFRβ* in developing neocortex have diverged considerably during human and mouse evolution, despite retaining amino acid sequences that are ~85% identical between the species.

We tested the requirement of *PDGFD*–*PDGFRβ* signalling for hRG proliferation in GW17.5 human neocortical slice cultures, screening four chemical inhibitors of *PDGFRβ* signalling (Sutent, tivozanib, imatinib and CP673451). Three out of four *PDGFRβ* inhibitors reduced the percentage of SOX2⁺ progenitors (RG) that incorporated 5-bromodeoxyuridine (BrdU) over 2 days in slice culture (Extended Data Fig. 9). For replication we focused on CP673451, which exhibits the greatest selectivity for *PDGFRβ* over other receptors²¹ and caused the greatest reduction

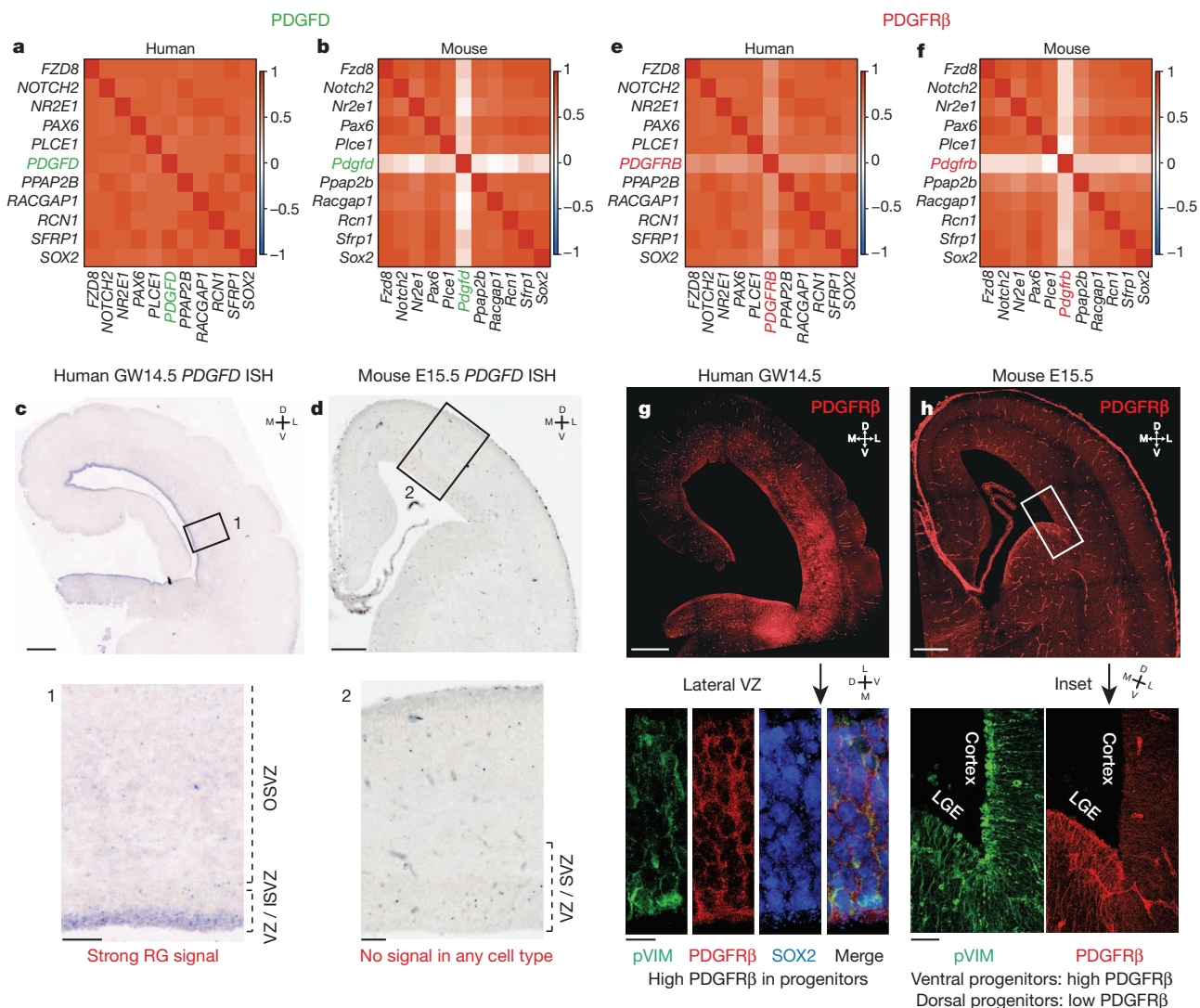


Figure 3 | PDGFD and PDGFRβ are expressed in human, but not mouse, dorsal RG. Consensus Pearson correlations among expression levels of *PDGFD* (a, human), *Pdgfd* (b, mouse), *PDGFRB* (e, human), or *Pdgfrb* (f, mouse) and 10 genes with the highest RG PR. **c, d**, ISH of *PDGFD* in human (c; scale bar, 1 mm; inset scale bar, 200 μm) and mouse neocortex (d; scale bar, 200 μm; inset scale bar, 50 μm). **g**, Immunostaining of *PDGFRβ* in human

neocortex (scale bar, 1 mm). **h**, Immunostaining of *PDGFRβ* in mouse neocortex (comparison with pVIM in the lateral ganglionic eminence (LGE); scale bar, 250 μm; inset scale bar, 50 μm).

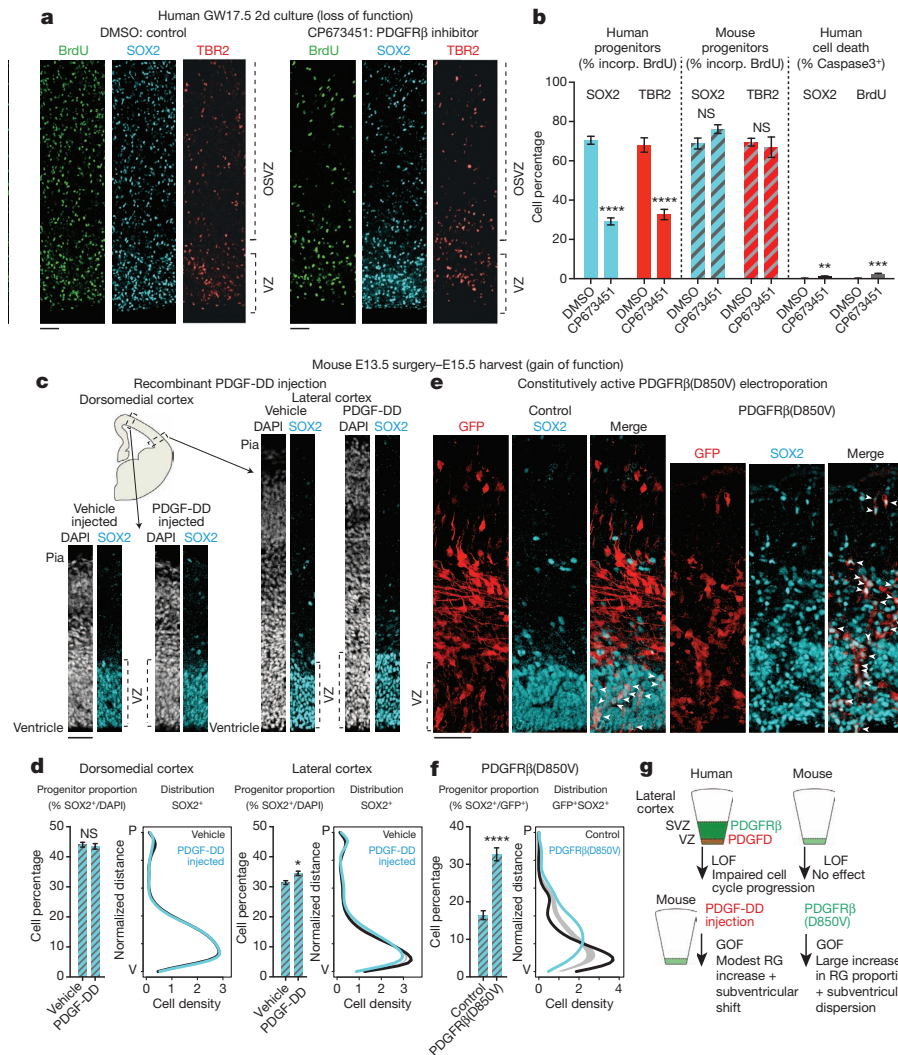


Figure 4 | PDGFD-PDGFR β signalling is necessary for normal cell cycle progression of neocortical RG in humans and sufficient to promote RG identity in mice. **a**, GW17.5 human neocortical slice cultures were treated with BrdU and DMSO (control) or an inhibitor of PDGFR β signalling (CP673451). Scale bar, 50 μ m. The same experiment was performed in E13.5 mouse neocortical slice cultures (slices from at least three individuals/litters per species). **b**, RG (intermediate progenitor) proliferation was quantified as the fraction of SOX2⁺ (TBR2⁺) cells that incorporated BrdU after 48 h. RG slice counts: human ($n = 18$ (DMSO) versus $n = 17$ (CP673451)); mouse ($n = 13$ (DMSO) versus $n = 11$ (CP673451)). Intermediate progenitor slice counts: human ($n = 12$ (DMSO) versus $n = 10$ (CP673451)); mouse ($n = 11$ (DMSO) versus $n = 9$ (CP673451)). Cell death was quantified in human slices as the fraction of SOX2⁺ or BrdU⁺ cells that co-stained for cleaved-caspase-3 ($n = 6$ (DMSO) versus $n = 7$ (CP673451)). **c**, *In utero* intraventricular injection of recombinant human PDGF-DD protein (mouse E13.5-E15.5). Brain tissue was stained for SOX2 and DAPI. Scale bar, 50 μ m. **d**, Quantification of data from

c in dorsomedial and lateral cortex (at least $n = 3$ slices per embryo from five litters/experiments (lateral: $n = 49$ vehicle; $n = 47$ PDGF-DD; dorsomedial: $n = 45$ vehicle; $n = 39$ PDGF-DD)). The distribution of RG in the cortex (from ventricle to pia) was quantified; grey band delineates 95% confidence interval for test of equal univariate densities ($n = 10,000$ permutations). **e**, *In utero* electroporation of constitutively active PDGFR β (D850V)²² (mouse E13.5-E15.5). Cortex was stained for SOX2; white arrowheads indicate co-labelling with electroporated GFP cells (quantified in **f**: at least $n = 3$ slices per embryo from two litters; $n = 15$ (control), $n = 18$ (PDGFR β (D850V))); scale bar, 50 μ m). Note disrupted epithelial structure of VZ. Error bars indicate mean \pm s.e.m. Statistical significance for treatment was determined by ANOVA of multiple linear regression after controlling for individual (**b**) or litter (**d**, **f**) (NS (not significant), $P > 0.05$; $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, $****P \leq 0.0001$). **g**, Schematic summarizing experimental manipulations and results. GOF, gain-of-function; LOF, loss-of-function.

in SOX2⁺BrdU⁺ cells among tested inhibitors (Extended Data Fig. 9). PDGFR β inhibition by CP673451 in GW17.5 human neocortical slice cultures reduced the number of RG and intermediate progenitors that incorporated BrdU by $>50\%$, affecting progenitors in the VZ and SVZ (Fig. 4a, b). The percentage of progenitor cells co-staining with cleaved caspase-3, an apoptosis marker, was slightly elevated by CP673451, but sufficiently low to attribute reduced BrdU incorporation to cell cycle dysregulation rather than cell death (Fig. 4b). Furthermore, CP673451 treatment of E13.5 mouse cortical slice cultures did not decrease BrdU incorporation or the cycling proportion (Ki67⁺) of RG or intermediate progenitor populations over multiple time points (Fig. 4b and Extended Data Fig. 9). These results indicate that PDGFR β signalling is required

for hRG but not mRG to progress through the cell cycle and expand at a normal rate.

We next investigated whether introducing PDGFD into embryonic mouse cortex, where it is normally absent, could promote mRG proliferation. We injected recombinant PDGF-DD protein into the lateral ventricles of E13.5 mouse embryos, bypassing the need for PDGFD to be generated and dimerized *in vivo*, and analysed the number and spatial distribution of SOX2⁺ progenitors at E15.5 in lateral cortex. Relative to vehicle, PDGF-DD increased the proportion of RG (SOX2⁺DAPI⁺) in lateral cortex by $\sim 10\%$ (Fig. 4c, d). Furthermore, PDGF-DD induced a modest subventricular shift in the distribution of mRG in the developing cortical wall (Fig. 4d). These effects were not observed in

cortex, where PDGFR β was not detected in RG (Fig. 4c, d and Extended Data Fig. 8).

To test whether mRG can respond to PDGFR β activation, we ectopically expressed two forms of constitutively active PDGFR β (PDGFR β (D850V) (homologous to human D849V (ref. 22)) or the TEL–PDGFR β (ref. 23) fusion protein) in mRG by *in utero* electroporation at E13.5. By E15.5, expression of PDGFR β (D850V) approximately doubled the proportion of SOX2⁺ or Ki67⁺ progenitors among electroporated (GFP⁺) cells and markedly dispersed SOX2⁺ progenitors in the basal direction (Fig. 4e, f and Extended Data Fig. 9). Similar but less marked effects were observed after TEL–PDGFR β electroporation (Extended Data Fig. 9). These results indicate that PDGFR β signalling in mRG can function analogously to its known role as an oncogenic pathway that promotes proliferation and epithelial–mesenchymal transition¹⁹. We therefore propose that physiological levels of PDGFR–PDGFR β signalling in hRG may contribute to the proliferation and subventricular dispersion of neural progenitors that characterize OSVZ formation^{4,5}.

Mouse studies have demonstrated that the size and shape of cerebral cortex depend on precise regulation of molecular pathways controlling RG proliferation and differentiation^{1,2}. While many of these pathways are probably conserved in humans, mouse studies alone cannot reveal uniquely human aspects of cortical development^{24–29}. By analysing human tissue as a starting point, we found that PDGFR–PDGFR β signalling is required for normal RG proliferation in developing human but not mouse neocortex, and is sufficient to promote some ‘humanizing’ characteristics in mouse. Our analysis has also identified other genes that probably contribute to differences between human and mouse cortical development. *BMP7*, another secreted growth factor, was predicted and validated to be expressed by hRG but not mRG (Fig. 2d and Extended Data Fig. 6), raising the possibility that local production of growth factors by hRG may be necessary to support the expanded germinal region and progenitor heterogeneity of developing human neocortex. We anticipate that the analytical and experimental strategies described here will help determine the extent to which these and other pathways are shared among primates or uniquely required for human cortical development.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 April; accepted 16 October 2014.

1. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nature Rev. Neurosci.* **10**, 724–735 (2009).
2. Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011).
3. Smart, I. H. et al. Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. *Cereb. Cortex* **12**, 37–53 (2002).
4. Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
5. Fietz, S. A. et al. OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nature Neurosci.* **13**, 690–699 (2010).
6. Bergsten, E. et al. PDGF-D is a specific, protease-activated ligand for the PDGF β -receptor. *Nature Cell Biol.* **3**, 512–516 (2001).
7. LaRochelle, W. J. et al. PDGF-D, a new protease-activated growth factor. *Nature Cell Biol.* **3**, 517–521 (2001).
8. Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).
9. Oldham, M. C. In *The OMIcs: Applications in Neuroscience* Vol. 1 (ed. Coppola, G.) 85–113 (Oxford Univ. Press, 2014).
10. Oldham, M. C. et al. Functional organization of the transcriptome in human brain. *Nature Neurosci.* **11**, 1271–1282 (2008).
11. Workman, A. D. et al. Modeling transformations of neurodevelopmental sequences across mammalian species. *J. Neurosci.* **33**, 7368–7383 (2013).

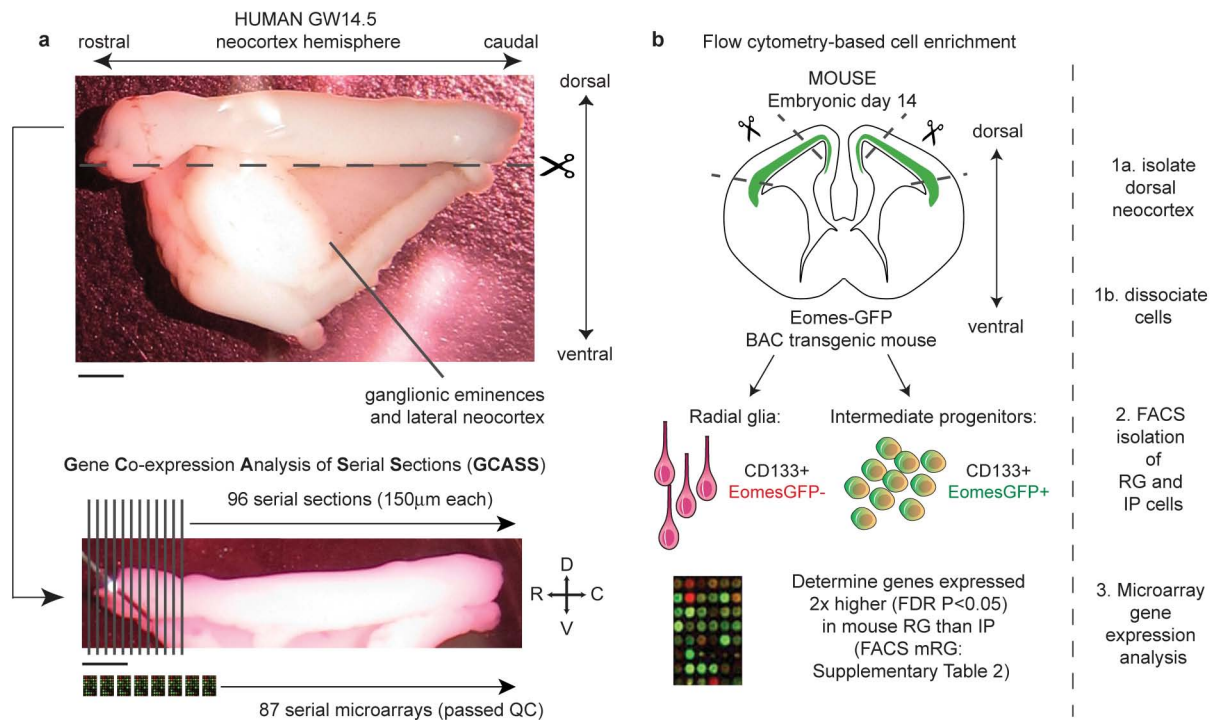
12. Fietz, S. A. et al. Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proc. Natl Acad. Sci. USA* **109**, 11836–11841 (2012).
13. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
14. Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLOS Comput. Biol.* **4**, e1000117 (2008).
15. Oldham, M. C., Horvath, S. & Geschwind, D. H. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc. Natl Acad. Sci. USA* **103**, 17973–17978 (2006).
16. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17 (2005).
17. Kawaguchi, A. et al. Single-cell gene profiling defines differential progenitor subclasses in mammalian neurogenesis. *Development* **135**, 3113–3124 (2008).
18. Pinto, L. et al. Prospective isolation of functionally distinct radial glial subtypes—lineage and transcriptome analysis. *Mol. Cell. Neurosci.* **38**, 15–42 (2008).
19. Wang, Z. et al. Emerging roles of PDGF-D signaling pathway in tumor development and progression. *Biochim. Biophys. Acta* **1806**, 122–130 (2010).
20. Diez-Roux, G. et al. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* **9**, e1000582 (2011).
21. Roberts, W. G. et al. Antiangiogenic and antitumor activity of a selective PDGFR tyrosine kinase inhibitor, CP-673,451. *Cancer Res.* **65**, 957–966 (2005).
22. Magnusson, P. U. et al. Platelet-derived growth factor receptor- β constitutive activity promotes angiogenesis *in vivo* and *in vitro*. *Arterioscler. Thromb. Vasc. Biol.* **27**, 2142–2149 (2007).
23. Golub, T. R., Barker, G. F., Lovett, M. & Gilliland, D. G. Fusion of PDGF receptor β to a novel *ets*-like gene, *tel*, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation. *Cell* **77**, 307–316 (1994).
24. McLean, C. Y. et al. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011).
25. Bae, B. I. et al. Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. *Science* **343**, 764–768 (2014).
26. Geschwind, D. H. & Rakic, P. Cortical evolution: judge the brain by its cover. *Neuron* **80**, 633–647 (2013).
27. Johnson, M. B. et al. Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009).
28. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
29. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff at San Francisco General Hospital Women's Options Center for their consideration in allowing access to donated human prenatal tissue. We thank J. DeYoung and the staff at the Southern California Genotyping Consortium at the University of California Los Angeles for microarray data generation. We are grateful to A. Holloway for her critical reading of the manuscript, and also thank W. Walantus, S. Wang, Y. Wang and other University of California San Francisco personnel for technical and administrative support. We thank C. Stiles and D. Rowitch for the TEL–PDGFR β construct. This work was supported by grants from the NIH, NINDS (A.R.K.), the Bernard Osher Foundation, a California Institute for Regenerative Medicine Predoctoral Fellowship for J.H.L. (TG2-01153), a Damon Runyon Foundation Postdoctoral Fellowship for A.A.P. (DRG-2013), and the University of California San Francisco Program for Breakthrough Biomedical Research, which is funded in part by the Sandler Foundation (M.C.O.). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the California Institute for Regenerative Medicine or any other agency of the State of California.

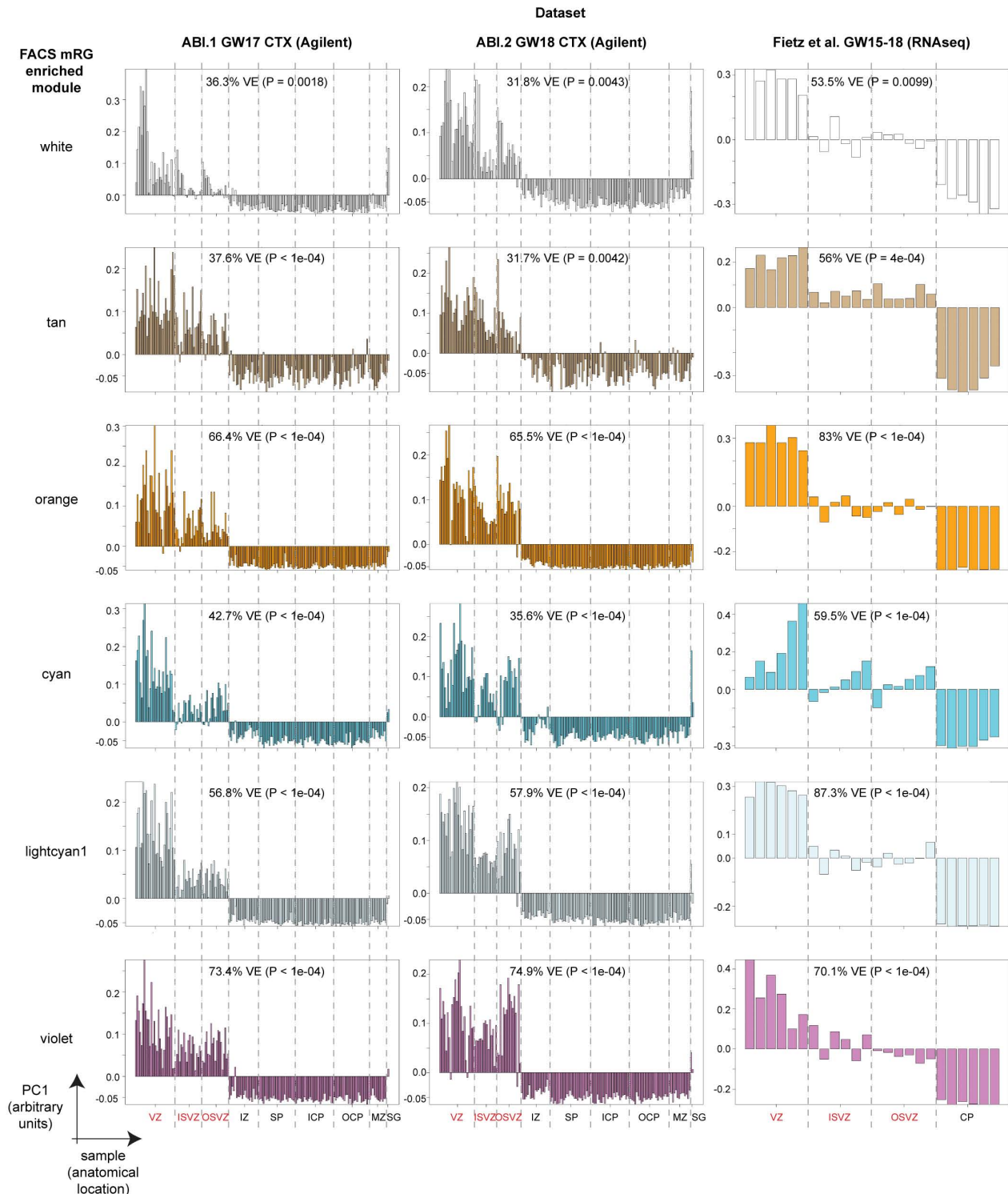
Author Contributions M.C.O. conceived the GCASS strategy and J.H.L. generated the GCASS data set. A.J. generated the FACS mRG data set. M.C.O. conceived, designed and performed the bioinformatic analyses. J.H.L., T.J.N. and A.A.P. designed and performed the experiments leading up to the prioritization of PDGFR as the focus of this study. T.J.N. performed the majority of the *in situ* hybridizations and the *in vivo* mouse experiments. J.H.L. performed the human and mouse slice culture experiments, as well as all of the immunostaining, imaging and image analysis in the study. M.C.O. and J.H.L. wrote the manuscript, which was edited by all the authors. M.C.O. and A.R.K. provided conceptual guidance at every stage of the project.

Author Information Microarray data from the GCASS dataset have been deposited in Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE62064. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C.O. (oldham@stemcell.ucsf.edu) or A.R.K. (kriegsteina@stemcell.ucsf.edu).



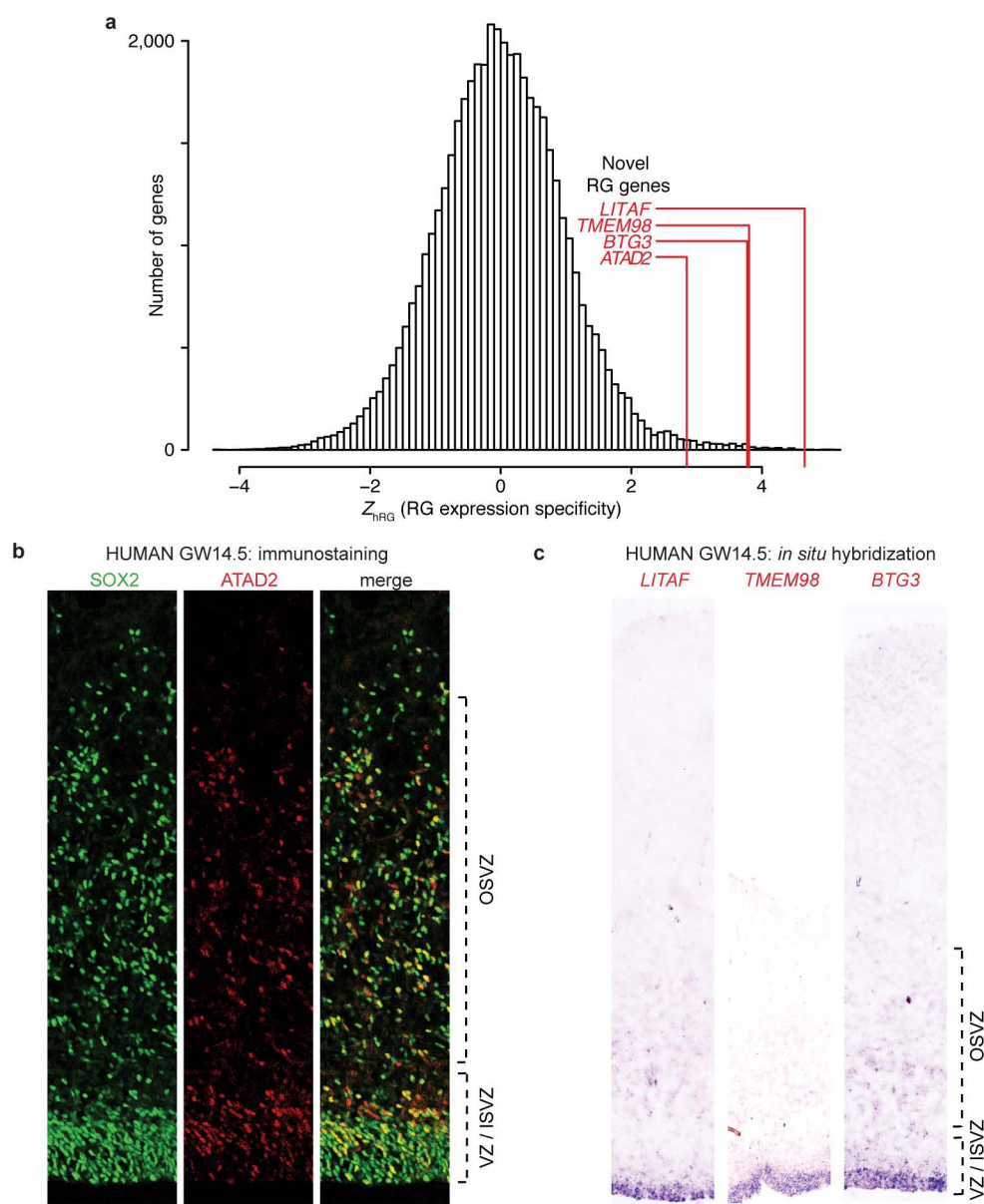
Extended Data Figure 1 | Human brain dissection for GCASS and schematic for generation of FACS mRG data set. **a**, Top: to generate the GCASS data set, an almost-intact prenatal human telencephalic hemisphere (GW14.5) was microdissected to separate the dorsal telencephalon from the ventral telencephalon (including medial and lateral ganglionic eminences). Bottom: the dorsal fragment was flash-frozen and serially sectioned (150 µm) for transcriptional profiling with Illumina HT-12 v4 Beadchip microarrays.

Scale bars, 2.5 mm. **b**, To generate the FACS mRG data set, dorsal neocortices of Eomes-GFP mouse embryos were microdissected, pooled ($n = 3$ litters, 5–8 pooled embryos per litter), dissociated and FACS-sorted according to the gating scheme depicted to isolate RG and intermediate progenitor cells. Transcriptional profiling of the resultant populations was performed using Illumina mouseRef-8 v1.0 Beadchip microarrays.



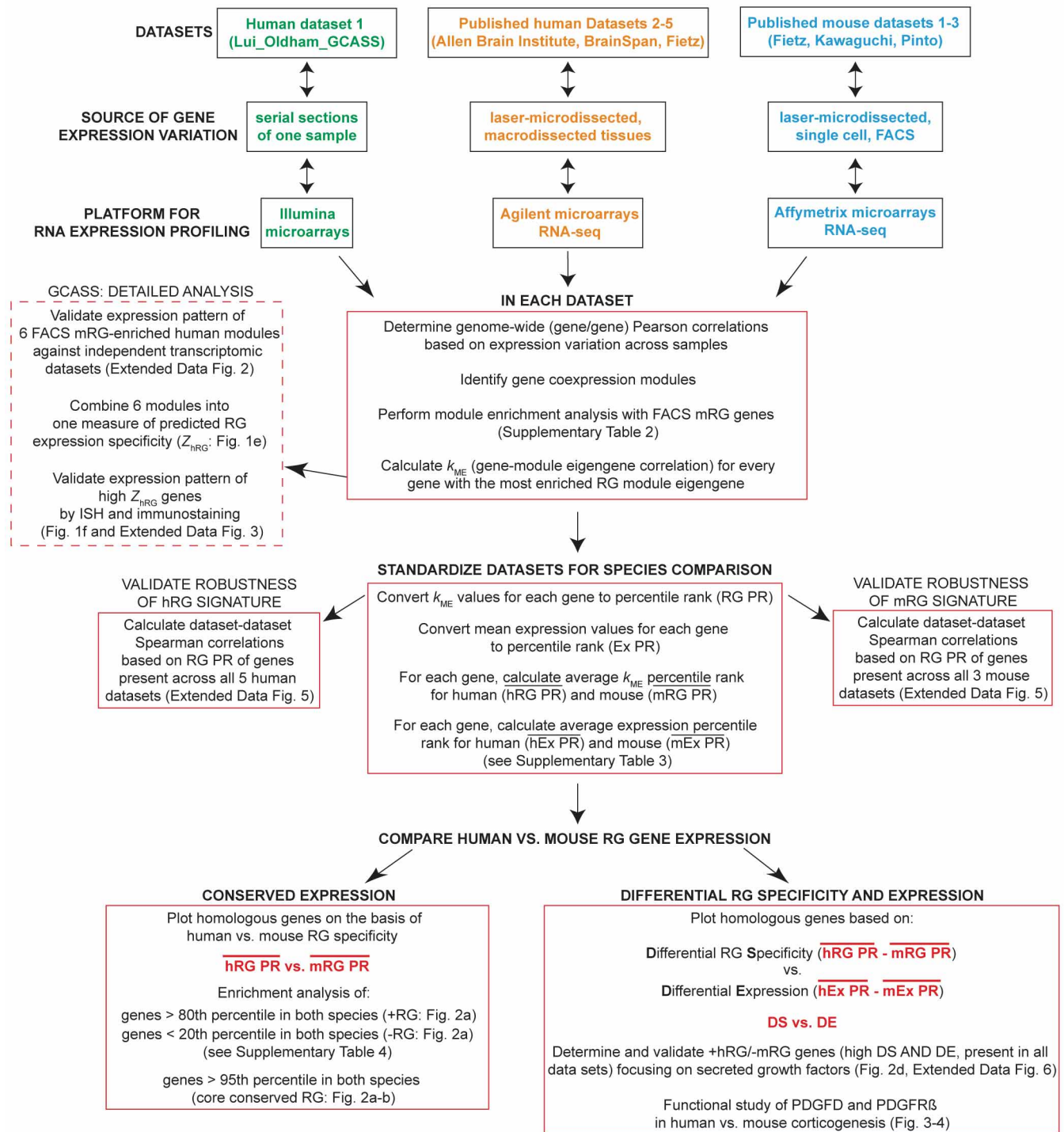
Extended Data Figure 2 | Genes comprising the six RG co-expression modules identified by GCASS are expressed in a manner consistent with the known distribution of RG in developing human neocortex. Six candidate hRG gene co-expression modules (Fig. 1d) were superimposed on three independent gene expression data sets generated from laser-microdissected samples from prenatal human cortex: ABI.1 (ref. 13) (GW17), ABI.2 (ref. 13) (GW18), and Fietz *et al.* (ref. 12) (GW15–18) (as listed in Extended Data Table 1). The characteristic expression patterns of the superimposed modules were summarized by singular value decomposition; the first principal component (PC1) for each module in each data set is shown. In all cases,

PC1 revealed substantially higher expression levels for these genes in germinal zones (VZ, ISVZ, and OSVZ, highlighted in red) versus non-germinal zones (IZ, SP, ICP, OCP, CP, MZ, SG). Permutation analysis indicated that the per cent variance explained (VE) by PC1 of each superimposed module was significantly greater than expected by chance ($n = 10,000$ permutations). CP, cortical plate; CTX, cortex; GW, gestational week; ICP, inner cortical plate; ISVZ, inner subventricular zone; IZ, intermediate zone; MZ, marginal zone; OCP, outer cortical plate; OSVZ, outer subventricular zone; SG, subpial granular layer; SP, subplate; VZ, ventricular zone.



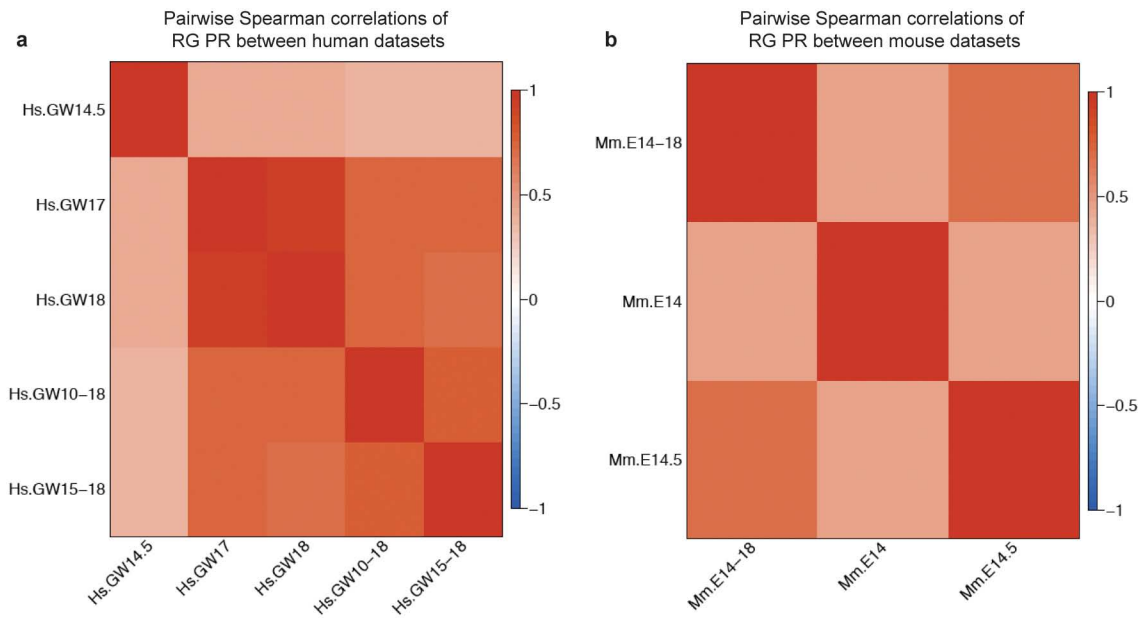
Extended Data Figure 3 | GCASS successfully predicts novel markers of neocortical hRG. **a**, Genome-wide distribution of predicted GW14.5 neocortical hRG expression specificity (Z_{hRG}). Red lines indicate predicted RG genes (validated in **b**, **c**). **b**, **c**, Immunostaining and *in situ* hybridization in GW14.5 human neocortex confirms RG expression specificity for novel

candidate markers predicted in **a** (**b**, scale bar, 50 μ m; **c**, scale bar, 100 μ m). Analysed tissue sections were independent from the sample used for microarray analysis (Fig. 1a). ISVZ, inner subventricular zone; OSVZ, outer subventricular zone; VZ, ventricular zone.



Extended Data Figure 4 | Workflow of bioinformatic procedures and experimental rationale for the entire study. The bioinformatic component of this study sought to identify a homologous gene co-expression signature for human and mouse RG that is robust across multiple sampling strategies/technology platforms and can be normalized to facilitate

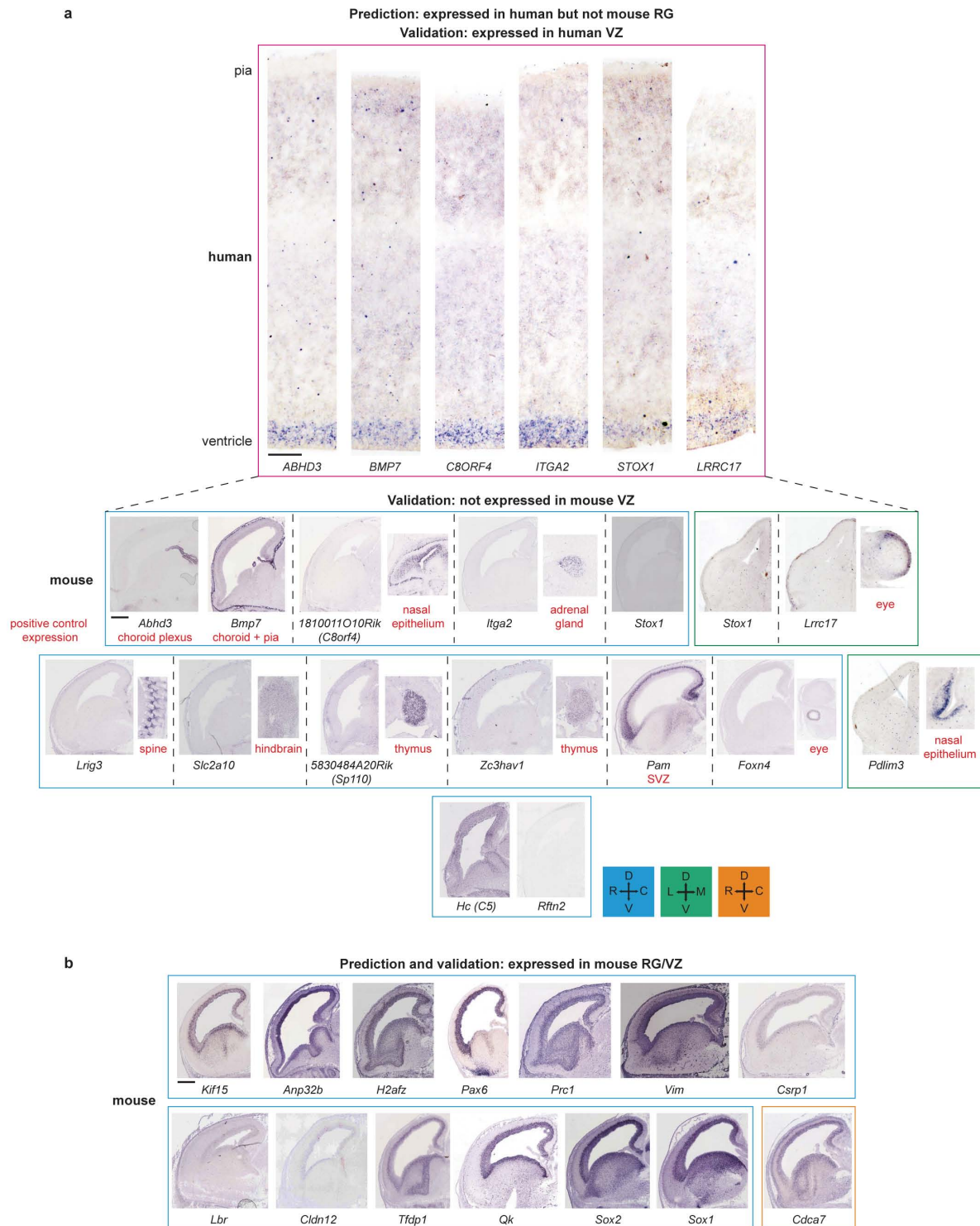
comparisons within and between species. This pipeline illustrates the steps that were taken to identify, integrate and compare RG gene co-expression signatures in eight transcriptomic data sets generated from prenatal human and mouse neocortex.



Extended Data Figure 5 | Genome-wide predictions of expression specificity for hRG and mRG are robust across independent data sets.

a, b, Heat maps of Spearman correlation coefficients for predicted RG expression specificity (RG PR) over 10,929 genes present in all five human data sets (**a**) and 10,649 genes present in all three mouse data sets (**b**) (as indicated in

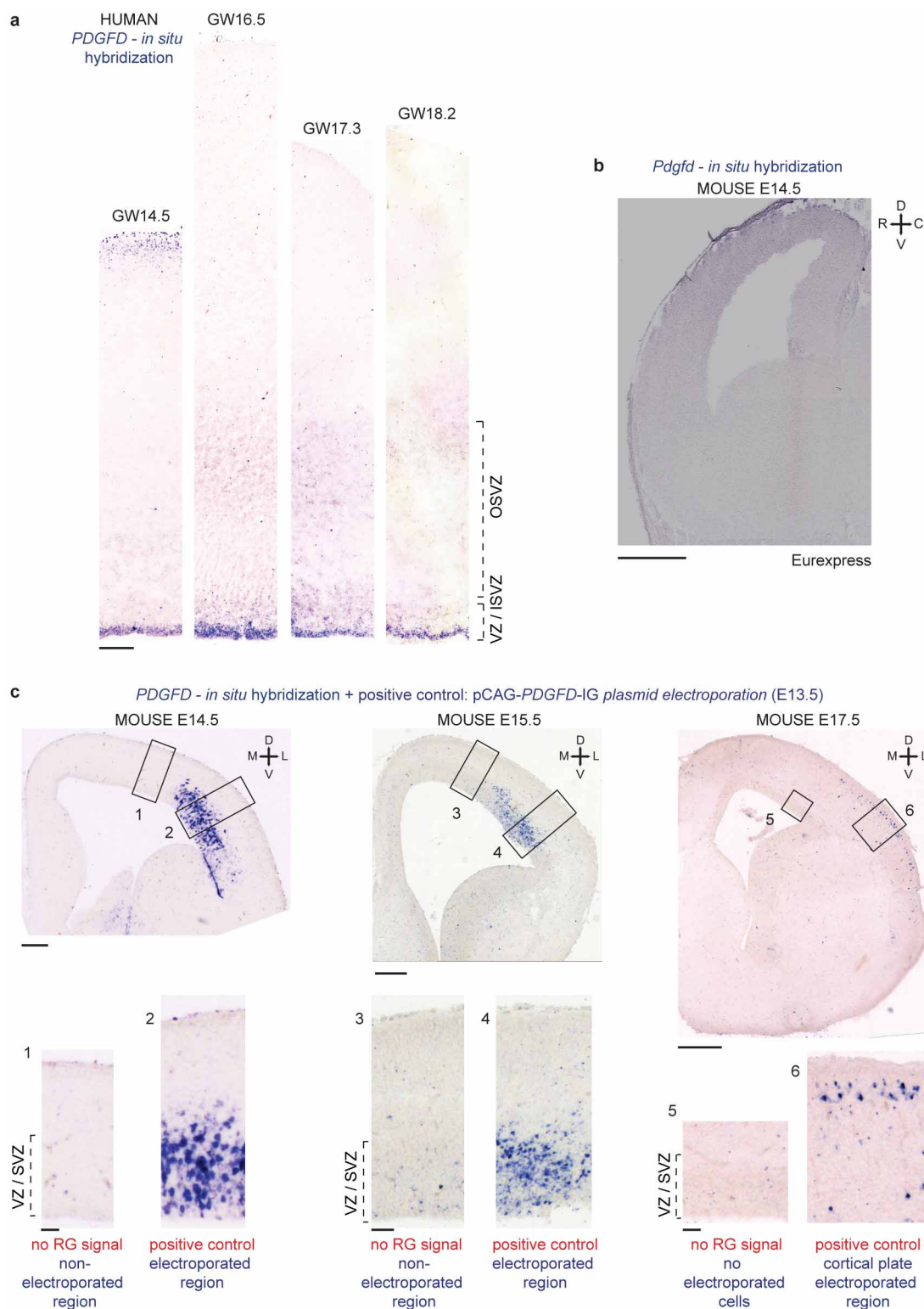
columns BE and BI in Supplementary Table 3). Data sets are denoted by the sample ages listed in Extended Data Table 1, although factors besides age also contribute to the observed correlations (for example, choice of technology platform, sample preparation strategy). E, embryonic; GW, gestational week.



Extended Data Figure 6 | *In situ* hybridization (ISH) validates predicted presence or absence of gene expression in hRG and mRG. **a**, Pink box: human *in situ* probes for six genes predicted to be expressed by hRG but not mRG were generated and hybridized in GW15 human neocortical tissue to validate predicted hRG expression (human scale bar, 200 μ m). Blue box (Eurexpress²⁰: <http://www.eurexpress.org/ee/>): *in situ* hybridizations for 13 genes predicted to be expressed by hRG but not mRG reveal no expression by mRG in E14.5 mouse cortex. Green box: mouse *in situ* probes for three genes

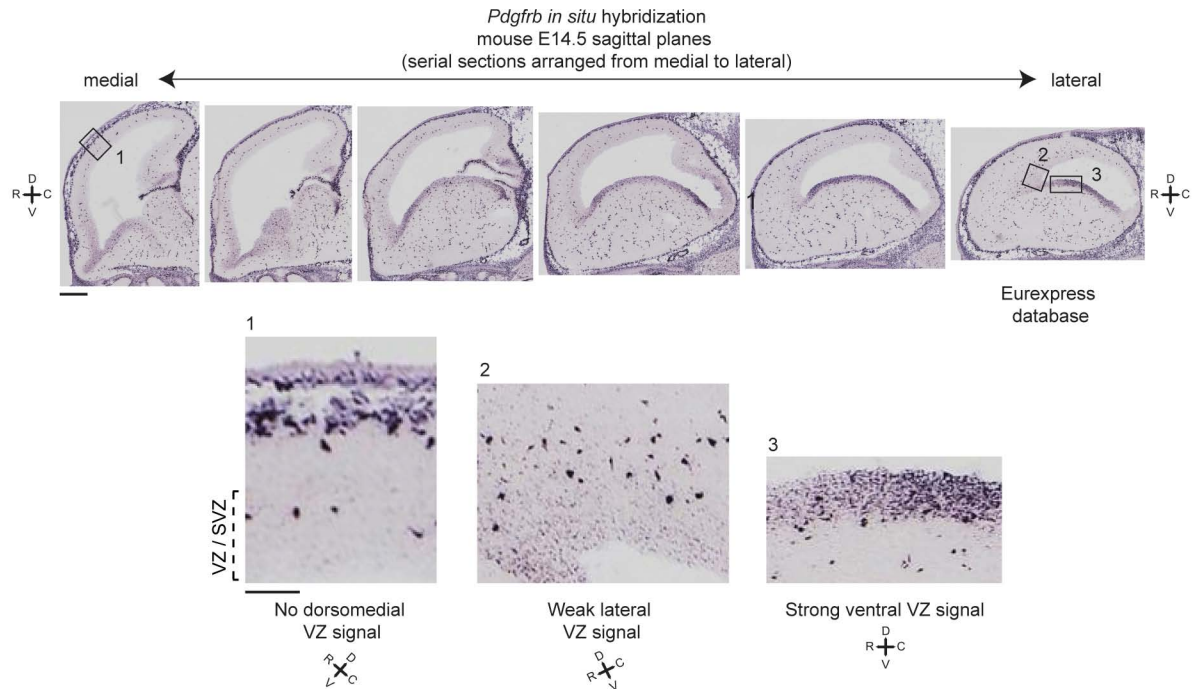
predicted to be expressed by hRG but not mRG were generated and revealed no expression by mRG (E13.5). Positive control expression in cells other than RG are labelled in red. **b**, Expression patterns of genes predicted to be expressed by mRG (that is, those with the highest mRG PR values in Supplementary Table 3) are shown as further validation (blue (E14.5, Eurexpress²⁰: <http://www.eurexpress.org/ee/>); orange (E14.5, GenePaint: <http://www.genepaint.org>); mouse scale bars, \sim 500 μ m). One other gene in the top 15, *Cks2*, is not shown, but has been validated previously³⁰.

30. Ajioka, I., Maeda, T. & Nakajima, K. Identification of ventricular-side-enriched molecules regulated in a stage-dependent manner during cerebral cortical development. *Eur. J. Neurosci.* **23**, 296–308 (2006).



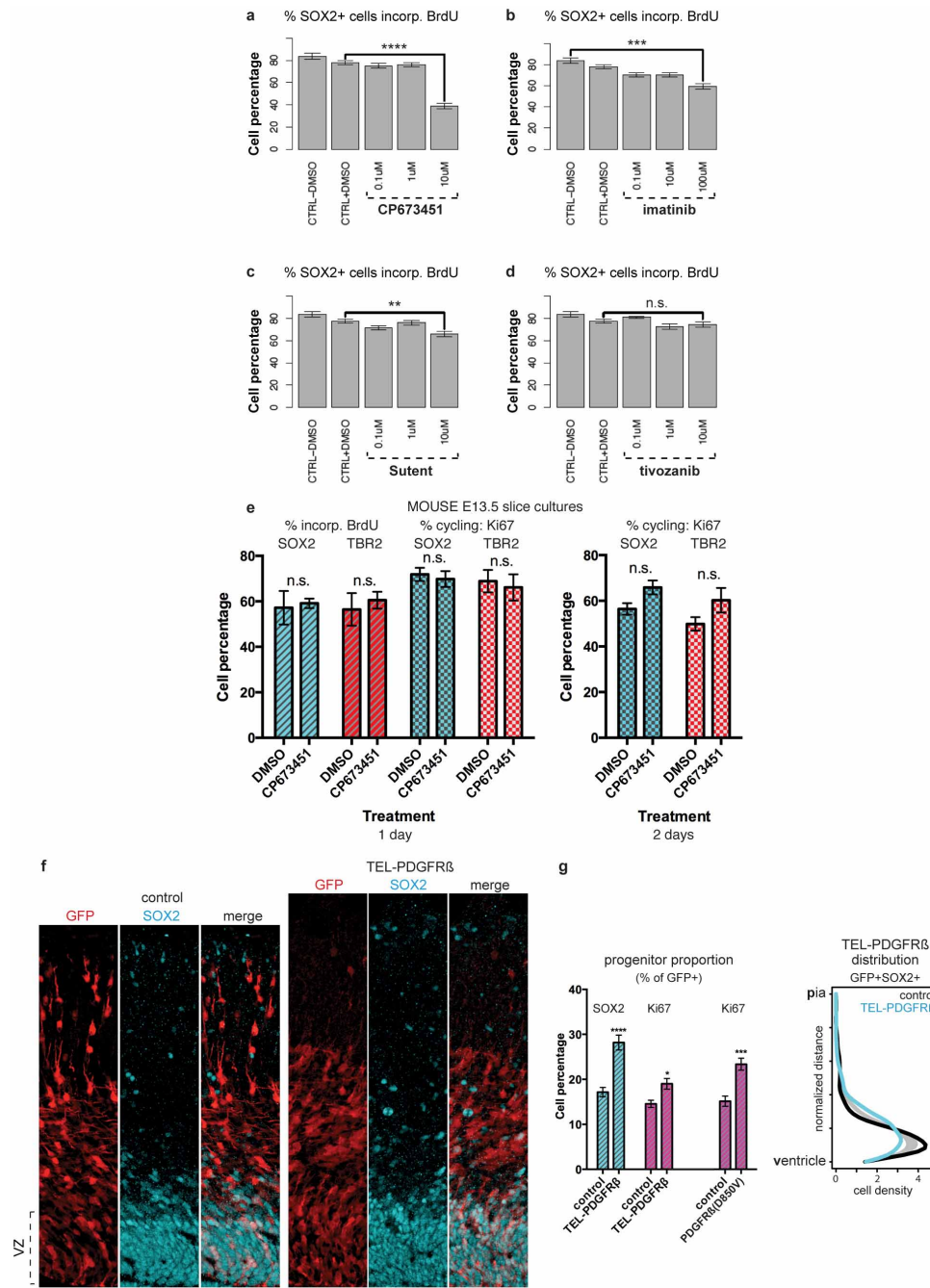
Extended Data Figure 7 | PDGFD is expressed by neocortical RG during neurogenesis in humans but not mice. **a**, *In situ* hybridization of *PDGFD* in GW14.5, 16.5, 17.3 and 18.2 human neocortex demonstrates consistent expression in RG across multiple ages (scale bar, 200 μ m). **b**, *In situ* hybridization of *Pdgfd* in E14.5 mouse (Eurexpress²⁰; <http://www.eurexpress.org/ee/>) demonstrates lack of expression (scale bar, ~500 μ m). **c**, To demonstrate the lack of *Pdgfd* expression in mouse neocortex across multiple

ages, a pCAG-*PDGFD*-IG expression plasmid was electroporated into the mouse VZ at E13.5 as an internal positive control and harvested at E14.5, E15.5 or E17.5. At E14.5 and E15.5, *Pdgfd* (blue signal) is seen only in the electroporated region in the ventricular zone (scale bar, 200 μ m; inset scale bar, 50 μ m). At E17.5, *Pdgfd* is in the cortical plate and not in the ventricular zone or anywhere else (scale bar, 500 μ m; inset scale bar, 50 μ m).



Extended Data Figure 8 | *Pdgfrb* is strongly expressed by ventral RG and weakly expressed by lateral RG in mice. *In situ* hybridization of *Pdgfrb* in sagittal sections through the mouse forebrain (E14.5) across a medial–lateral axis (Eurexpress²⁰: <http://www.eurexpress.org/ee/>) demonstrates progenitor expression in the ventral germinal regions. This expression extends into

the dorsal cortex in the lateral aspect of the brain, but is not widespread. In contrast, no progenitor expression is detected in dorsomedial cortex (scale bar, 500 μ m; inset scale bar, 100 μ m). Expression is also detected in the pia and vascular pericytes.



Extended Data Figure 9 | Manipulation of PDGFR β signalling in human and mouse neocortex. **a–d**, Chemical blockade of PDGFR β signalling in cultured slices of GW14.5 human neocortex impairs RG cell cycle progression. Four pharmacological inhibitors of PDGFR β signalling were screened at different concentrations to determine their effects on RG proliferation in cultured slices of GW17.5 human neocortex (2 days). Slices were treated with BrdU for the duration of the experiment and RG proliferation was quantified as the fraction of SOX2 $^{+}$ cells that incorporated BrdU after treatment with inhibitor or vehicle. Statistical significance was assessed with the Wilcoxon rank sum test using the wilcox.test R function with default settings. Images are derived from ≥ 3 slices in each condition. Control + DMSO, $n = 18$; control no DMSO, $n = 9$; CP673451 all concentrations, $n = 9$; Sutent all concentrations, $n = 6$; imatinib (0.1 μ M, 10 μ M), $n = 9$ (100 μ M), $n = 6$; tivozanib (1 μ M, 10 μ M), $n = 9$ (0.1 μ M), $n = 6$. Significance indicated by: n.s. $P > 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, $****P \leq 0.0001$. **e**, Slice cultures of E13.5 mouse neocortex were treated with BrdU and DMSO (control) or a pharmacological inhibitor of PDGFR β signalling (CP673451) for 1 or 2 days (slices from at least three independent litters). RG (SOX2 $^{+}$) or intermediate progenitor (TBR2 $^{+}$) cell proliferation was assessed as the fraction of each population that

incorporated BrdU or was Ki67 $^{+}$ (1d: $n = 10$ (DMSO) versus $n = 9$ (CP673451); 2d: $n = 11$ (DMSO) versus $n = 9$ (CP673451)). This experiment serves as a negative control to compare with the human. **f**, Ectopic PDGFR β signalling promotes RG identity in E13.5 mouse neocortex. *In utero* electroporation of constitutively active TEL-PDGFR β (ref. 23) was compared with control (mouse E13.5–E15.5) and assessed for the proportion and distribution of SOX2 $^{+}$ RG cells or Ki67 $^{+}$ progenitors (out of GFP $^{+}$) in the cortical wall (quantified in **g**; at least $n = 3$ slices per embryo from two independent litters, $n = 15$ (control); $n = 18$ (TEL-PDGFR β) or (PDGFR β (D850V)); scale bar, 50 μ m). Ki67 $^{+}$ GFP $^{+}$ cell quantification following PDGFR β (D850V) 22 electroporation was performed in a similar fashion and is also shown. The spatial distributions of RG (GFP $^{+}$ SOX2 $^{+}$) in the cortical wall were assessed by quantitative image analysis (spanning ventricle to pia). The grey band delineates a 95% confidence interval for a test of equal univariate densities based on 10,000 permutations. All error bars represent mean \pm s.e.m. Statistical significance for the effects of treatment was calculated by ANOVA of multiple linear regression while controlling for individual (**e**) and litter (**f**) variability (significance indicated by: n.s. $P > 0.05$, $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, $****P \leq 0.0001$).

Extended Data Table 1 | Characteristics of additional gene expression data sets analysed in the present study

Dataset	Platform	Species	CTX prep	Age	Samples	mRG enrichment <i>P</i> value	OR
ABI.1 ¹³	Agilent	Human	LMD	GW17	179	2.6e-39	7.2 [5.5-9.5]
ABI.2 ¹³	Agilent	Human	LMD	GW18	168	9.7e-35	7.5 [5.6-10.0]
BrainSpan	RNAseq	Human	Macro	GW10-18	107	2.0e-47	9.2 [6.9-12.1]
Fietz et al. ¹²	RNAseq	Human	LMD	GW15-18	24	1.3e-39	8.4 [6.3-11.1]
Fietz et al. ¹²	RNAseq	Mouse	LMD	E14.5	15	1.3e-73	12.5 [9.6-16.4]
Kawaguchi et al. ¹⁷	Affymetrix	Mouse	Single cell	E14	70	3.1e-34	59.0 [34.2-100.6]
Pinto et al. ¹⁸	Affymetrix	Mouse	FACS	E14-18	17	7.8e-46	20.7 [14.7-29.0]

Unsupervised gene co-expression analysis was performed for each data set using the same parameters, followed by unbiased enrichment analysis with the FACS mRG gene set (Supplementary Table 2 and Extended Data Fig. 1). For BrainSpan RNA-seq data see also (<http://www.brainspan.org/rnaseq/search/index.html>). 'mRG enrichment *P* value' denotes the *P* value (Fisher's exact test) for the module with the most significant enrichment in each data set. ABI: Allen Institute for Brain Science; CTX, cortex; E, embryonic; FACS, fluorescence-activated cell sorting; GW, gestational week; LMD, laser-microdissection; Macro, macrodissection; OR, odds ratio (95% confidence interval).

The participation of cortical amygdala in innate, odour-driven behaviour

Cory M. Root¹, Christine A. Denny^{2,3,4}, René Hen^{3,4,5} & Richard Axel¹

Innate behaviours are observed in naive animals without prior learning or experience, suggesting that the neural circuits that mediate these behaviours are genetically determined and stereotyped. The neural circuits that convey olfactory information from the sense organ to the cortical and subcortical olfactory centres have been anatomically defined^{1–3}, but the specific pathways responsible for innate responses to volatile odours have not been identified. Here we devise genetic strategies that demonstrate that a stereotyped neural circuit that transmits information from the olfactory bulb to cortical amygdala is necessary for innate aversive and appetitive behaviours. Moreover, we use the promoter of the activity-dependent gene *arc* to express the photosensitive ion channel, channelrhodopsin, in neurons of the cortical amygdala activated by odours that elicit innate behaviours. Optical activation of these neurons leads to appropriate behaviours that recapitulate the responses to innate odours. These data indicate that the cortical amygdala plays a critical role in generating innate odour-driven behaviours but do not preclude its participation in learned olfactory behaviours.

Odours can elicit an array of innate behaviours including feeding, mating, freezing or escape, responses essential for the survival and reproduction of the organism. Innate responses to odours can be mediated by either the vomeronasal or main olfactory system⁴. The vomeronasal organ recognizes non-volatile odorants, including the major urinary proteins⁵ and steroids⁶, that elicit innate responses via a circuit emanating from the accessory olfactory bulb. The main olfactory system recognizes volatile cues including 2,3,5-trimethyl-3-thiazoline (TMT) in fox secretions⁷ and trace amines in bobcat⁸ and mouse⁹ urine that elicit innate attraction and avoidance responses.

Olfactory perception is initiated by the recognition of odorants by a large repertoire of receptors in the sensory epithelium^{10–12}. Neurons expressing a given receptor are randomly distributed within zones of the epithelium but project with precision to two spatially invariant glomeruli in the olfactory bulb^{13–15}. Thus a transformation in the representation of olfactory information is apparent in the bulb where the dispersed population of active neurons in the sense organ is consolidated into a discrete spatial map of glomerular activity¹⁶. This invariant glomerular map in the bulb is transformed in the representations in higher olfactory centres. Anatomical tracing experiments reveal that the projections from the olfactory bulb to the cortical amygdala retain a topographic map with individual glomeruli projecting to broad but spatially invariant loci¹. In contrast, spatial order in the bulb is discarded in the piriform cortex; axons from individual glomeruli project diffusely to the piriform cortex without apparent spatial preference^{1–3}. The identification of a distributive pattern of projections to the piriform cortex and stereotyped projections to the cortical amygdala provides an anatomical substrate for the generation of learned and innate behaviours. We have devised behavioural assays and genetic strategies to identify the olfactory centres responsible for innate odour-driven behaviours.

Innate behaviours are often complex and comprise multiple components, but can be simplified by the design of assays that categorize behaviours by only a single axis, positive or negative valence. We therefore developed an open field behavioural assay to determine whether a given odour elicits attraction or avoidance as a measure of innate odour valence. A symmetrical chamber was constructed with four quadrants allowing independent airflow into each of the quadrants with a vacuum in the centre. In the absence of odour, mice explored the chamber without bias for any quadrant (Extended Data Fig. 1f). However, the addition of TMT to one quadrant resulted in significant avoidance (Fig. 1a and Extended Data Fig. 1).

We describe the valence of behavioural response to an odour by defining a performance index (see legend to Fig. 1), such that a negative value indicates avoidance and a positive value indicates attraction. In this behavioural assay, TMT revealed a performance index of -83 ± 5.0 ($n = 5$). TMT not only elicited avoidance, but also sitting in distant quadrants, resulting in a marked decrease in locomotor activity, which may be a proxy for freezing (Extended Data Fig. 1e, h). This behavioural assay was used to examine the response of mice to several odorants. We observed that a small group of odorants elicited reliable approach or avoidance behaviour, but most odours elicited no significant behavioural response (Fig. 1b). For example, peanut oil and 2-phenylethanol, a component of rose oil, elicited attraction (performance index 50 ± 4.1 and 49 ± 4.8 ($n = 5$) respectively), whereas the odorants TMT, isopentylamine and 4-methylthiazole elicited aversion (performance index -83 ± 5.0 , -81 ± 4 and -59 ± 8.1 ($n = 4$) respectively). Most odours were neutral in this assay and revealed values of performance index near zero, ranging from -7 to 9 . Therefore, this robust assay is able to classify the innate valence of odours in terms of approach and avoidance responses.

We have used this behavioural assay in concert with optogenetic silencing to identify a higher olfactory centre necessary for innate olfactory-driven behaviours. The mitral and tufted cells of the olfactory bulb project axons to at least five brain regions¹: piriform cortex, cortical amygdala, entorhinal cortex, accessory olfactory nucleus and olfactory tubercle. We expressed the light-activated chloride pump, halorhodopsin¹⁷, in neurons of the olfactory bulb to permit optical silencing of the individual projections to these olfactory centres and examined the consequences on innate odour-driven behaviours.

Bilateral injection of an adeno-associated virus (AAV) encoding halorhodopsin fused to enhanced yellow fluorescent protein (AAV-syneNpHR3.0-eYFP) resulted in expression in the vast majority of mitral cells in the olfactory bulb ($80\% \pm 7$ (s.d.); Fig. 2a, b). Intrinsic neurons within the bulb also expressed halorhodopsin-eYFP, but mitral and tufted cells provide the only feed-forward excitatory output to cortical centres. In initial experiments, we asked whether silencing of neurons within the olfactory bulb suppressed the innate behaviour elicited by TMT. After bilateral infection, both olfactory bulbs were illuminated by introducing optical fibres coupled to a 561 nm laser above the dorsal surface of

¹Department of Neuroscience and the Howard Hughes Medical Institute, College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA. ²Department of Biological Sciences, New York State Psychiatric Institute, New York, New York 10032, USA. ³Department of Neuroscience and Psychiatry, New York State Psychiatric Institute, New York, New York 10032, USA. ⁴Division of Integrative Neuroscience, New York State Psychiatric Institute, New York, New York 10032, USA. ⁵Department of Pharmacology, Columbia University, New York State Psychiatric Institute, New York, New York 10032, USA.

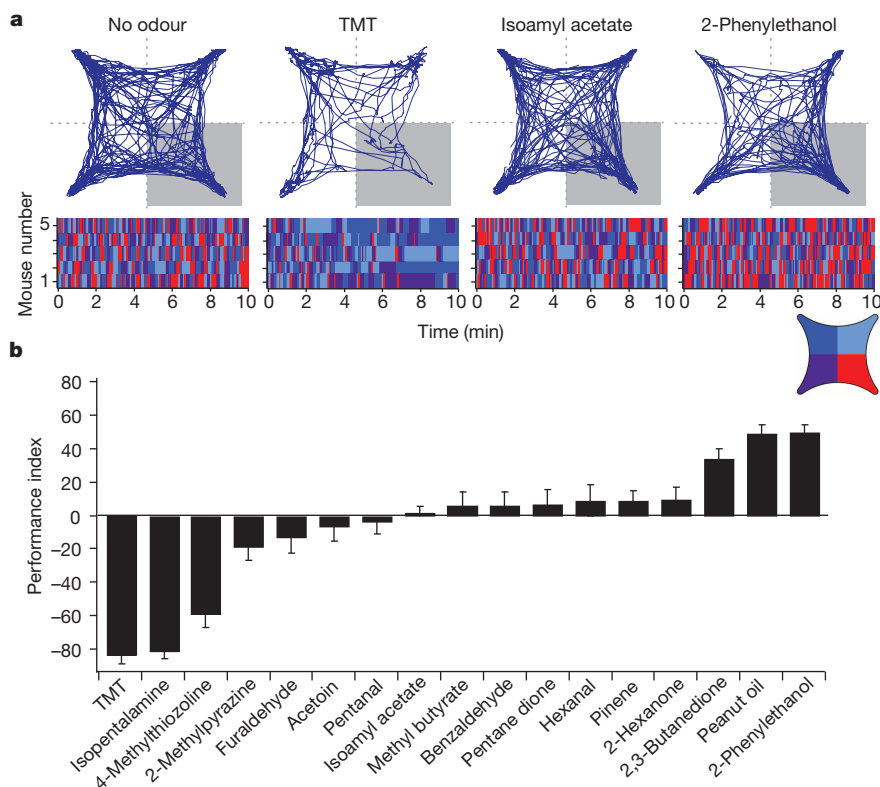


Figure 1 | Behavioural assay for innate responses to odour. An open field, four-quadrant behavioural chamber was used to measure the response to odour delivered in only one quadrant. **a**, The trajectory of a representative mouse is plotted for a 10 minute period in the absence of odour or after the addition of odour to the lower right quadrant. The raster plots below the trajectory graphs represent quadrant occupancy over time (x axis) for each of five different animals. The four colours represent occupancy in each of the four quadrants. Odour was delivered to the lower right quadrant (red). **b**, The average response to an array of odorants is quantified by a performance index that represents the percentage difference from chance occupancy in the lower right quadrant (performance index = $(P - 25) / 0.25$, where P is the percentage time in the lower right quadrant). One-way analysis of variance (ANOVA) test, $P < 0.001$; $n = 4-8$ for each odour.

the bulb. Previous experiments suggest that the glomeruli responsible for innate behaviour are restricted to the dorsal bulb⁷. Mice were then introduced into the behavioural chamber in which TMT was present within one quadrant and innate aversion was examined in the absence or presence of 561 nm illumination to silence bulbar activity. We observed that light-activated silencing of the olfactory bulb significantly suppressed the aversion to TMT with a reduction in the performance index from -82 ± 3.8 to -18 ± 2.9 ($n = 3$). These experiments demonstrate that these viral injections into the bulb result in halorhodopsin expression in neurons essential for innate aversive behaviour.

We then used optical silencing of the individual targets¹⁸ of the olfactory bulb to identify the olfactory centres necessary to elicit innate behaviour. Photostimulation of the cortical amygdala (Extended Data Fig. 2), for example, in mice expressing halorhodopsin in bulbar neurons should selectively silence bulbar input to this brain structure without affecting input to other olfactory centres. Mice were coupled to a 561 nm laser and placed in the four-field behavioural assay with TMT in a single quadrant. Each of eleven mice exhibited a striking reduction in the avoidance of the TMT quadrant upon bilateral illumination of the cortical amygdala (performance index -65 ± 3.4 without photostimulation, and -7.9 ± 8.4 upon optical silencing) (Fig. 2d). Further, silencing bulbar input significantly reduced the freezing behaviour, as evidenced by decreased bouts of inactivity (Extended Data Fig. 3). The inhibition of innate avoidance observed upon optical silencing of the cortical amygdala was reversible; robust avoidance re-emerged upon cessation of light-induced silencing (performance index -74 ± 5.7). In control animals not injected with virus, aversive behaviour was not impaired by photostimulation (Fig. 2f). We determined the efficacy of silencing upon illumination of bulbar axons by analysing *c-fos* activity. We observe a 70% reduction in the frequency of cells activated by odour in cortical amygdala but not olfactory tubercle or piriform cortex (Extended Data Fig. 4). These results demonstrate that axonal silencing is sufficient and suggests that antidromic hyperpolarization is not responsible for the suppression of behaviour.

We next determined whether input from the olfactory bulb to cortical amygdala is also required for innate attraction. We observed that light-induced silencing of olfactory input to cortical amygdala eliminated

attraction to 2-phenylethanol (performance index without photostimulation, 48 ± 3.2 ; performance index with photostimulation, -8.3 ± 2.1) (Fig. 2e, g and Extended Data Fig. 3b, d), but attraction resumed upon cessation of optical silencing. Thus, input from the olfactory bulb to the cortical amygdala is required to elicit innate responses to both appetitive and aversive odours.

We also asked whether suppression of olfactory input to either piriform cortex or olfactory tubercle, two additional targets of the bulb, would suppress innate, odour-driven behaviours. Bilateral illumination of axons from the olfactory bulb to either the tubercle or the piriform failed to suppress the innate aversion to TMT (performance index -61 ± 3.4 and -67 ± 4.0 respectively, $n = 7$ and 8) (Fig. 2f). Conclusions from these experiments, however, must be tempered by the fact that olfactory inputs to both the tubercle and piriform are extensive and the area of illumination might have been inadequate to suppress axonal input significantly. Nonetheless, this result provides an additional control for antidromic suppression of mitral and tufted cell activity that would result in suppression of innate behaviour independent of the site of optical silencing.

These observations predict that silencing of neurons intrinsic to the cortical amygdala should also inhibit innate behaviour. Halorhodopsin was therefore expressed in the neurons of the cortical amygdala after bilateral injections of AAV-eNpHR3.0-eYFP and bilaterally implanted optical fibres (Extended Data Fig. 5). In each of five mice, we observed a significant decrease in aversion of the TMT quadrant in the four-field behavioural assay (performance index before stimulation, -70 ± 4.8 ; after photostimulation, -15 ± 7.1) (Fig. 2f). Thus, light-induced silencing of the olfactory bulb, the bulbar projections to cortical amygdala or the neurons of the cortical amygdala dramatically impairs innate aversion to TMT.

Our data suggest that distinct cell populations that reside within the cortical amygdala are capable of eliciting innate responses to either appetitive or aversive odours. We next identified and manipulated the activity of these neurons by exploiting the promoter of the activity-dependent gene *arc*¹⁹, to drive the expression of the light-activated cation channel channelrhodopsin²⁰. AAV encoding a Cre-dependent channelrhodopsin fused to the fluorescent protein eYFP (AAV-ef1 α -DIO-ChR2-eYFP)

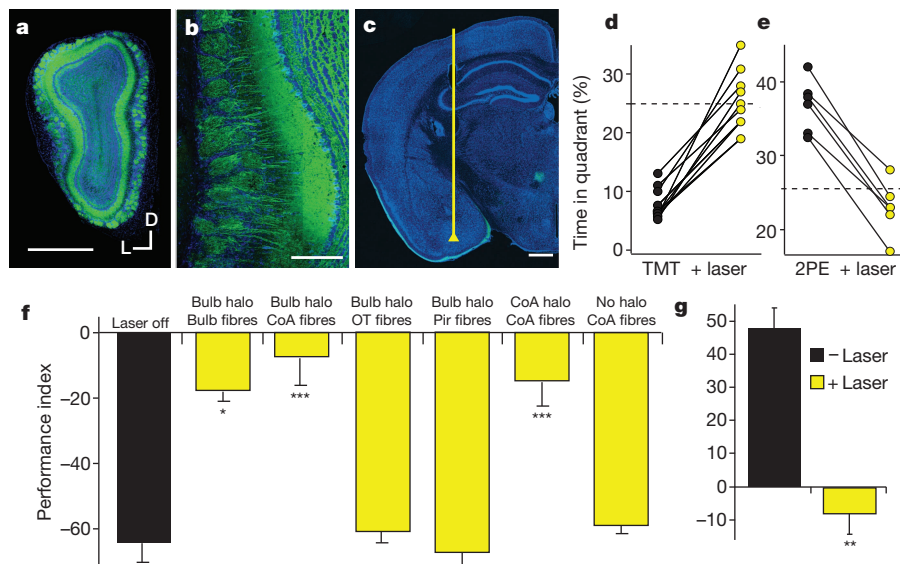


Figure 2 | The projections from the olfactory bulb to cortical amygdala are required for innate aversion and attraction to odours. **a**, Coronal section of a mouse olfactory bulb injected with AAV5-eNpHR3.0-eYFP. D, dorsal; L, lateral. **b**, Magnified view of an olfactory bulb showing eNpHR3.0-eYFP expression in mitral cells. **c**, Coronal section depicting the placement of an optical fibre in cortical amygdala, above the axonal output from the olfactory bulb. Scale bars, 500 μ m (**a**, **c**) and 100 μ m (**b**). **d**, **e**, Mice were optically coupled to a yellow laser and tested in the behavioural assay for the response to TMT (**d**) or 2-phenylethanol (2PE) (**e**) with and without laser stimulation. The charts show the percentage time each animal spent in the odour quadrant in the absence or presence of photoactivation. **f**, The mean performance index for mice exposed to TMT. The black bar indicates the average response of all mice

to TMT in the absence of photoactivation; the yellow bars indicate responses to TMT with photoactivation for different experimental animals. Bulb halo and CoA halo describe mice with halorhodopsin expression in the olfactory bulb and cortical amygdala, respectively. Optical fibres were placed above the bulb ($n = 3$), cortical amygdala (CoA, $n = 11$), olfactory tubercle (OT, $n = 7$) or in piriform cortex (Pir, $n = 8$) as denoted below the site of injection. The last two bars on the right side have either halorhodopsin in the neurons of cortical amygdala ($n = 5$), or receive no viral injection ($n = 4$), and fibres implanted into cortical amygdala. **g**, The mean performance index for mice exposed to 2-phenylethanol in the absence and presence of photoactivation ($n = 6$). **f**, **g**, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, paired t -test comparing performance index with and without laser for each group; error bars, s.e.m.

was injected into the cortical amygdala of mice harbouring a transgene in which the *arc* promoter drives the expression of the tamoxifen-sensitive Cre-recombinase (CreER^{T2}) (Fig. 3a). In this ArcCreER^{T2} mouse²¹, neuronal activation should result in the expression of CreER^{T2}. In the presence of tamoxifen, activated Cre will effect the recombination between the loxP sites of AAV-ef1 α -DIO-ChR2, resulting in the irreversible expression of ChR2-eYFP. The fusion of Cre with the tamoxifen-sensitive oestrogen receptor allows temporal control of the recombination activity. Thus administration of tamoxifen followed by exposure to odour should result in the expression of ChR2-eYFP in the neurons activated by the odour (Fig. 3b, c), permitting us to mark and manipulate the activity of these neural populations.

In initial experiments, we determined whether ChR2-eYFP is faithfully expressed in neurons that respond to specific odours in the cortical amygdala. ChR2-eYFP should be expressed by neurons responsive to TMT and subsequent re-exposure of the mice to TMT should result in the acute expression of endogenous *arc* in ChR2-eYFP positive neurons. This temporal separation permitted us to identify populations of neurons activated by sequential stimuli in the same animal and to determine whether the expression of ChR2-eYFP is a faithful reporter of endogenous *arc* activity. We observed that $80\% \pm 5.9$ (s.d.) of the ChR2-eYFP+ neurons also expressed endogenous *arc* ($n = 4$). About 50% of the neurons expressing endogenous *arc* also expressed ChR2-eYFP (Extended Data Fig. 6a, c). These results indicate that the expression of ChR2-eYFP is a faithful reporter of endogenous *arc* activity, identifying a population of neurons responsive to TMT.

We next asked whether odours that elicit innate behaviours of different valence activate different populations of neurons in the cortical amygdala. AAV encoding Cre-dependent ChR2 was injected into the cortical amygdala of ArcCreER^{T2} mice. After tamoxifen treatment, the mice were sequentially exposed to the attractive odour 2-phenylethanol and the aversive odour TMT. In this experimental design, the neurons

responsive to 2-phenylethanol will express ChR2-eYFP, whereas the neurons responsive to TMT will express the endogenous *arc*. Only $4\% \pm 2.1$ (s.d.) of the ChR2-eYFP+ neurons activated by 2-phenylethanol also expressed endogenous *arc*, activated by TMT ($n = 4$) (Extended Data Fig. 6b, c). These data suggest that odours that elicit appetitive and aversive innate behaviours activate different populations of neurons in the cortical amygdala that can be faithfully identified and manipulated after AAV infection of an ArcCreER^{T2} mouse.

The ArcCreER^{T2} mice were used to determine whether the cortical amygdala is sufficient to elicit a response that recapitulates the behaviour observed with odour. We therefore injected AAV encoding Cre-dependent ChR2-eYFP into the cortical amygdala of ArcCreER^{T2} mice and optical fibres were implanted unilaterally above the injection site. Mice were treated with tamoxifen and then exposed to odour. We observed ChR2 expression in cortical amygdala and neighbouring areas of piriform cortex and medial amygdala. However, c-fos staining demonstrated that optical activation is restricted to cortical amygdala as a consequence of fibre placement (Extended Data Fig. 7b). Behavioural assays were performed in the four-field arena, but odour was not placed in one quadrant: rather, the entrance to a single quadrant triggered pulsed laser stimulation via a custom closed-loop computer program.

Photoactivation of the cortical amygdala of mice expressing ChR2 after TMT exposure resulted in avoidance of the optically stimulated quadrant (performance index -49 ± 4.6 , $n = 6$) (Fig. 3d, g) and increased freezing as evidenced by increased immobility (Extended Data Fig. 8). In contrast, mice expressing ChR2 in neurons responsive to the innately attractive odour 2-phenylethanol exhibited significant attraction to the quadrant in which the mice received optical stimulation (performance index 42 ± 6.4 , $n = 4$) (Fig. 3e, g). Mice expressing channelrhodopsin in neurons responsive to the neutral odour isoamyl acetate did not exhibit any discernible behavioural response upon optical stimulation (performance index 5.6 ± 8.6 , $n = 6$) and explored each quadrant equally

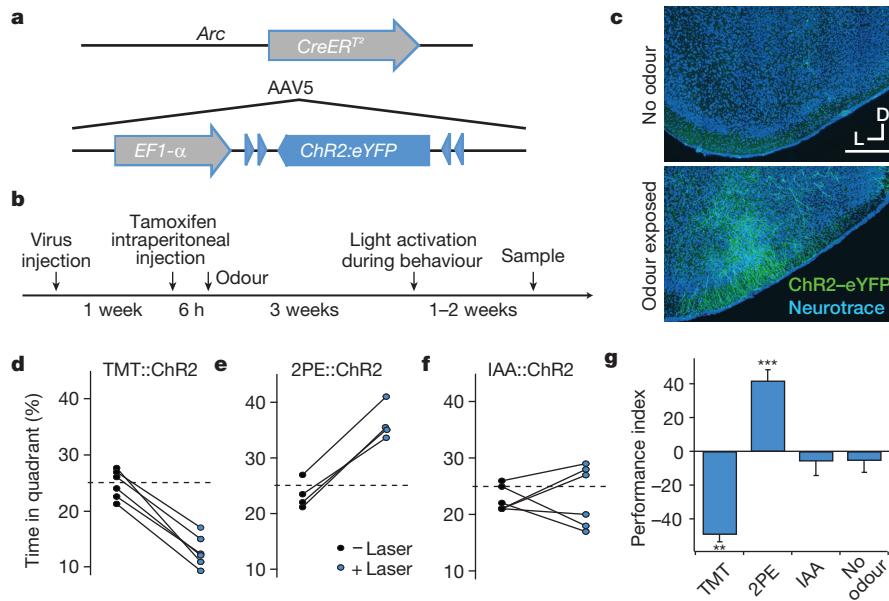


Figure 3 | Activation of odour-responsive neurons within cortical amygdala is sufficient to recapitulate behavioural responses. **a**, The genetic strategy used to express ChR2 in odour-responsive neurons. The tamoxifen-sensitive Cre recombinase CreERT² was expressed under the control of the promoter of the activity-dependent gene, *arc*, in a transgenic mouse. The gene encoding Cre-dependent ChR2-eYFP was introduced into the cortical amygdala by infection with AAV5. **b**, The timeline for experimental manipulations. The animal was sampled upon termination of behavioural testing. **c**, Representative images showing the expression of ChR2-eYFP in mice that received tamoxifen injection followed by exposure to either the odorant

isoamyl acetate (bottom) or no odour as a control (top). Scale bar, 300 μ m. **d–f**, Mice with odour-driven channelrhodopsin expression were tested in the open field assay where they received pulsed photoactivation upon entrance into one quadrant. The percentage time each animal spent in the lower right quadrant in the absence and presence of pulsed photoactivation in mice with neurons activated by TMT (**d**), 2-phenylethanol (**e**) and isoamyl acetate (IAA) (**f**). **g**, The average performance index for mice receiving photostimulation of neurons activated by TMT, 2-phenylethanol, isoamyl acetate or no odour, respectively, from left to right ($n = 3–6$). ** $P < 0.01$, *** $P < 0.001$, t -test comparing performance index with and without laser; error bars, s.e.m.

(Fig. 3f, g). In control mice, treatment with tamoxifen without odour exposure resulted in ChR2 expression in a very small subpopulation of neurons. These mice did not exhibit any behavioural bias upon photostimulation within a single quadrant (performance index -5.2 ± 7.1 , $n = 3$) (Fig. 3g). These experiments demonstrate that odours that elicit innate behaviours of different valence activate different populations of neurons within the cortical amygdala. Moreover, photoactivation of these two distinct populations of neurons is sufficient to elicit an appropriate behavioural response. Thus, these neural representations reflect an essential component in a determined neural circuit wired to elicit an innate behavioural response to odours.

The presence of spatially stereotyped projections from individual glomeruli to the cortical amygdala¹ suggest that the neural representations activated by aversive and attractive odours may be anatomically segregated. We therefore examined the spatial distribution of neurons activated by odour after AAV infection of the ArcCreERT² mouse (Fig. 4a, b and Extended Data Fig. 9). We observed that the ChR2-eYFP⁺ neurons, activated by the appetitive odours, 2-phenylethanol and peanut oil, were most abundant in the caudal third of the posterolateral cortical amygdala. Sparse labelling was observed in more anterior regions. In contrast, the aversive odours TMT, isopentylamine and 4-methylthiozoline activated neurons distributed throughout the cortical amygdala. Significant labelling was observed in more anterior regions that were not activated by odours that elicited appetitive behaviours. Neutral odours such as isoamyl acetate also failed to activate neurons in the anterior region and revealed labelling in a pattern not dissimilar from that observed with appetitive odours.

A given odour, even those that elicit innate behaviours, may activate as many as 100 glomeruli, and the inputs from these glomeruli may activate neurons distributed diffusely across the cortical amygdala. Neurons in the most anterior portion of the cortical amygdala, however, may only be activated by odours that elicit innate aversive behaviours. Our data suggest a model in which aversive odours activate a small subset

of glomeruli that project to anterior cortical amygdala and a larger set of glomeruli that project more broadly throughout the cortical amygdala. Regions responsible for appetitive behaviours may reside in more posterior domains; however, a locus that is uniquely responsive to appetitive odours is not easily distinguishable in our analyses. Neurons in the cortical amygdala are also activated by neutral odours that do not elicit an apparent behavioural response. We do not know the functional significance of the neural representation of neutral odours. These neurons may also reside in segregated domains that encode innate features of an odour, including subtle behaviours or odour quality, that are not discernible in our assays. Alternatively, the neurons in cortical amygdala responsive to neutral odour may participate in the adaptive response to odours, a function thought to engage the piriform cortex²².

The response to most odours is not innate but adaptive. Most odours have predictive value only over the lifetime of the organism and acquire meaning through learning, a function suited to the unstructured representation in piriform cortex^{23–25}. Exogenous activation of an arbitrarily chosen ensemble of piriform neurons can result in appetitive, aversive and social behaviours only after associative learning²². Therefore, odour representations in piriform cortex are afforded behavioural significance upon experience. Our data indicate that the cortical amygdala plays a critical role in the generation of innate odour-driven behaviours but do not preclude the participation of cortical amygdala in learned olfactory behaviours. Cortical amygdala may cooperate with piriform cortex to impose meaning on a random ensemble of active neurons in piriform. One major target of piriform output is the cortical amygdala²⁶. Piriform connections in the cortical amygdala may therefore exploit the distinct subpopulations of neurons in cortical amygdala capable of eliciting innate behaviours to generate learned responses to odour. In one model, direct connections from the olfactory bulb to cortical amygdala may elicit innate odour-driven behaviours, whereas indirect connections from the bulb through piriform to cortical amygdala may elicit learned responses to odour. This indirect, trisynaptic pathway may facilitate

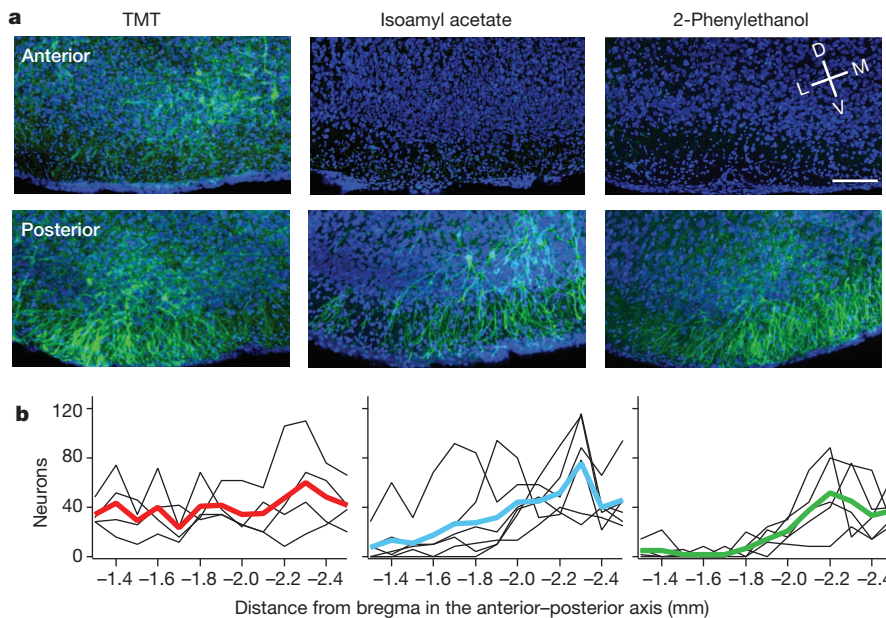


Figure 4 | The spatial distribution of neurons responsive to different odours within the cortical amygdala. ArcCreER^{T2} mice were administered tamoxifen and exposed to one of three different odours and then consecutive, serial, coronal sections were collected throughout the cortical amygdala region. **a**, Images reveal odour-driven ChR2-eYFP expression in representative images of anterior (top) and posterior (bottom) cortical amygdala for each odour. Images of the ventral brain region are magnified and cropped to show only the posterolateral cortical amygdala, and sections are displayed at -1.4 and -2.2 mm relative to bregma. Scale bar, $100\ \mu\text{m}$. D, dorsal; L, lateral; M, medial; V, ventral. **b**, The number of neurons counted per $100\ \mu\text{m}$ throughout the posterolateral cortical amygdala is displayed for each odour. The number for each animal is indicated with a thin black line; the heavier coloured line shows the mean for the odour group ($n = 4-6$).

learning-dependent synaptic changes necessary to impart valence on a random ensemble of neurons. The convergence of direct monosynaptic input and indirect polysynaptic input, capable of more complex computations from the same brain structure, is a common motif both in vertebrate²⁷ and invertebrate²⁸ brains.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 April; accepted 26 September 2014.

Published online 5 November 2014.

- Sosulski, D. L., Bloom, M. L., Cutforth, T., Axel, R. & Datta, S. R. Distinct representations of olfactory information in different cortical centres. *Nature* **472**, 213–216 (2011).
- Miyamichi, K. *et al.* Cortical representations of olfactory input by trans-synaptic tracing. *Nature* **472**, 191–196 (2011).
- Ghosh, S. *et al.* Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature* **472**, 217–220 (2011).
- Stowers, L. & Logan, D. W. Olfactory mechanisms of stereotyped behavior: on the scent of specialized circuits. *Curr. Opin. Neurobiol.* **20**, 274–280 (2010).
- Chamero, P. *et al.* Identification of protein pheromones that promote aggressive behaviour. *Nature* **450**, 899–902 (2007).
- Nodari, F. *et al.* Sulfated steroids as natural ligands of mouse pheromone-sensing neurons. *J. Neurosci.* **28**, 6407–6418 (2008).
- Kobayakawa, K. *et al.* Innate versus learned odour processing in the mouse olfactory bulb. *Nature* **450**, 503–508 (2007).
- Dewan, A., Pacifico, R., Zhan, R., Rinberg, D. & Bozza, T. Non-redundant coding of aversive odours in the main olfactory pathway. *Nature* **497**, 486–489 (2013).
- Li, Q. *et al.* Synchronous evolution of an odor biosynthesis pathway and behavioral response. *Curr. Biol.* **23**, 11–20 (2013).
- Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175–187 (1991).
- Godfrey, P. A., Malnic, B. & Buck, L. B. The mouse olfactory receptor gene family. *Proc. Natl Acad. Sci. USA* **101**, 2156–2161 (2004).
- Zhang, X. & Firestein, S. The olfactory receptor gene superfamily of the mouse. *Nature Neurosci.* **5**, 124–133 (2002).
- Ressler, K. J., Sullivan, S. L. & Buck, L. B. Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* **79**, 1245–1255 (1994).
- Mombaerts, P. *et al.* Visualizing an olfactory sensory map. *Cell* **87**, 675–686 (1996).
- Vassar, R. *et al.* Topographic organization of sensory projections to the olfactory bulb. *Cell* **79**, 981–991 (1994).
- Rubin, B. D. & Katz, L. C. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* **23**, 499–511 (1999).
- Gradinaru, V., Thompson, K. R. & Deisseroth, K. eNpHR: a Natronomonas halorhodopsin enhanced for optogenetic applications. *Brain Cell Biol.* **36**, 129–139 (2008).
- Tye, K. M. *et al.* Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature* **471**, 358–362 (2011).
- Link, W. *et al.* Somatodendritic expression of an immediate early gene is regulated by synaptic activity. *Proc. Natl Acad. Sci. USA* **92**, 5734–5738 (1995).
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
- Denny, C. A. *et al.* Hippocampal memory traces are differentially modulated by experience, time, and adult neurogenesis. *Neuron* **83**, 189–201 (2014).
- Choi, G. B. *et al.* Driving opposing behaviors with ensembles of piriform neurons. *Cell* **146**, 1003–1014 (2011).
- Stettler, D. D. & Axel, R. Representations of odor in the piriform cortex. *Neuron* **63**, 854–864 (2009).
- Poo, C. & Isaacson, J. S. Odor representations in olfactory cortex: “sparse” coding, global inhibition, and oscillations. *Neuron* **62**, 850–861 (2009).
- Illig, K. R. & Haberly, L. B. Odor-evoked activity is spatially distributed in piriform cortex. *J. Comp. Neurol.* **457**, 361–373 (2003).
- Schwabe, K., Ebert, U. & Loscher, W. The central piriform cortex: anatomical connections and anticonvulsant effect of GABA elevation in the kindling model. *Neuroscience* **126**, 727–741 (2004).
- Tonegawa, S. & McHugh, T. J. The ins and outs of hippocampal circuits. *Neuron* **57**, 175–177 (2008).
- Tanaka, N. K., Tanimoto, H. & Ito, K. Neuronal assemblies of the *Drosophila* mushroom body. *J. Comp. Neurol.* **508**, 711–755 (2008).

Acknowledgements We thank L. Abbott, T. Jessell and D. Costantini for comments and reading the manuscript; B. Bader for assistance with experiments; M. Mendelsohn and N. Zabello for help with mice; P. Kisloff for assistance in preparation of the manuscript; and A. Nemes and M. Gutierrez for laboratory support. This work was supported by the Howard Hughes Medical Institute and the Mathers Foundation.

Author Contributions C.M.R. and R.A. conceived the project, participated in its development, wrote the manuscript and analysed data. C.M.R. performed all experiments. C.A.D. and R.H. conceived and generated the ArcCreER^{T2} transgenic mouse.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A. (ra27@columbia.edu).

METHODS

Behavioural assay. The four-field behaviour chamber was inspired by similar assays used in experiments with *Drosophila* (see refs 29–31). The custom behavioural chamber was machined from plastic materials with the help of the Columbia University machine shop on the Nevis campus. The chamber was an enclosed four-quadrant arena with airflow into the centre and a vacuum in the middle. Airflow at 150 ml min^{-1} was pumped into each quadrant via gas-mass flow controllers (Cole-Parmer). Airflow exited the chamber via a 1-inch outlet in the centre of the floor covered by steel mesh. The outlet was connected a vacuum line with a gas-mass flow controller set to 700 ml min^{-1} . The chamber had a 76 cm diameter, a corner arc of 1.9 cm radius and a chamber height of 7.6 cm. The floor and the walls were made out of polypropylene and the hinged ceiling was clear, $\frac{1}{2}$ -inch thick acrylic. Odour was applied by solenoid valves redirecting airflow through 100 ml glass bottles containing 1 μl of pure odorant on a small piece of Kimwipe. The chamber was housed in a dark environment and illuminated by infrared lights below the floor. A Basler A601FM camera (Edmund Optics) mounted above the chamber recorded the animals' behaviour. Custom software written in Labview (National Instruments) tracked the position of the mouse in real time at a rate of 4 Hz. The behaviour was further analysed offline using custom, unbiased scripts written in Igor Pro (Wavemetrics).

Mice were placed in the chamber for 25 min experiments and tested only once per day. The first 10 min served as a baseline test for bias within the arena. The first 2 minutes of data after the odour was introduced were excluded from the analysis to reduce variance without affecting the overall valence of the behavioural response. The symmetrical behaviour chamber was contained in a lightproof structure and illuminated by infrared lights. Therefore, there were no spatial cues available to the animals with respect to the room, and the odour was always delivered in the same quadrant. To control for any spatial bias, animals were always tested with a 10 min period of no odour. If an animal avoided any one quadrant more than 20% from chance, the experiment was terminated and the animal was tested again on another day. For the experiments in Fig. 1, this occurred in fewer than 10% of the trials and a given animal was never excluded from the data set. For the optogenetic experiments, mice were more inclined to sessions of inactivity resulting in bias for one area during the baseline period. Repeated baseline testing acclimated the animals to being tethered in the chamber, which typically permitted the experiment to proceed after one or two baseline sessions. About half of the animals required more than one baseline period before the experiment could continue. Of the 65 mice tested in the optogenetic experiments, only four were rejected because of unremitting problems with locomotor activity or spatial bias during the baseline period. Moreover, we observed that a given animal could be retested to the same odour at least five times on consecutive days and exhibited no experience-dependent bias during the baseline, and then exhibited reliable responses to the odour (data not shown). Therefore, mice did not exhibit any signs of conditioned place preference or learned responses to odour in this assay. Experiments in Fig. 1a–d were performed with the same set of animals, and those in Fig. 1e were obtained from overlapping groups of animals. Odours were tested in random order. A preliminary set of experiments revealed that, depending on the variance and magnitude of difference, a sample as small as four animals was sufficient to achieve significant findings.

Optogenetic behaviour experiments. Animals were only tested once per day, and the laser on/off sessions occurred at least 1 day apart. Half of the mice experienced odour with the laser on upon the first test, and the other half experienced the odour with the laser off first. Mice were implanted with custom-made fibre cannulas assembled as follows. Optical fibres (200 μm , 0.39 numerical aperture, Thorlabs) were epoxied to 2.5 mm stainless steel ferrules (Precision Fibre Products), and polished with a fibre optic polishing kit (Thorlabs) to achieve a minimum of 80% transmission. After surgical implantation, the ferrules protruding from an animal's head could be coupled to either a 100 mW 473 nm or 100 mW 561 nm laser (Shanghai Laser & Optics Century) via custom-made patch cords with either a single, or beam splitting, rotary joint (Doric Lenses) between the mouse and the laser. The fibre was introduced to the behavioural chamber through a 1 cm hole in the left side of the chamber. During the behavioural experiment, it was necessary to pull the slack of the fibre as the mouse traversed the chamber to prevent the mouse from getting tangled in the fibre. Therefore, a motorized pulley system was created to pull the fibre-coupled rotary joint along a vertical tract adjacent to the hole where the fibre entered the chamber. Custom closed-loop software written in Labview monitored the position of the mouse and drove the pulley system to adjust the height of the rotary joint to maintain the appropriate amount of slack. The blue laser was pulsed with 50 ms bins at 10 Hz, and there was a steep gradient from 1 to 10 Hz along the perimeter of the quadrant. The lasers were controlled by transistor–transistor logic (TTL) modulation from custom Labview software.

Light intensity. Fibres were implanted approximately 200–400 μm above layer II of the cortical amygdala, which should have illuminated an area with a radius no more than 600 μm . For halorhodopsin experiments, the 561 nm laser power was adjusted to produce an estimated 10 mW at the implanted fibre tip, which we

calculated as providing an irradiance of $9\text{--}22 \text{ mW mm}^{-2}$ at layer II of the cortical amygdala. Suppression in the olfactory bulb was performed by placing optical fibres bilaterally 0.5–1 mm above the centre of the dorsal surface of the bulb. We estimated the conical spread of light from the fibre tip illuminated an area of about 3.5 mm in diameter with at least 7 mW mm^{-2} . In these experiments, the fibre was introduced through a bilateral cannula positioned in the centre of the olfactory bulbs about 0.5 mm above the surface at coordinates relative to bregma (~ 5.0 anterior–posterior, 1.0 medial–lateral). The olfactory bulb was roughly 3 mm in length and 1–2 mm wide, so fibres placed above the centre would illuminate most of the dorsal surface. For channelrhodopsin experiments, the 473 nm laser was adjusted to an estimated $5\text{--}7 \text{ mW}$ at the implanted fibre tip, which we calculated as providing an irradiance of $7\text{--}18 \text{ mW mm}^{-2}$ at layer II of the cortical amygdala. A conservative estimate of the radius of illumination was calculated with trigonometry using the half angle of divergence for a multimode optical fibre: $\theta = \sin^{-1}(\text{NA}/n)$, where NA is the numerical aperture of the fibre (0.39) and n is the index of refraction of grey matter (1.36) (ref. 32). However, it has been suggested that, because of scattering in tissue, the lateral spread is quantitatively similar to the depth of forward light spread from the fibre tip³³. Thus, at 400 μm below the fibre tip, the radius of illumination would be 400 μm from the fibre edge. Power attenuation was calculated as described³⁴:

$$\frac{I_z}{I_{z0}} = \frac{\tilde{n}^2}{(S_z + 1)(z + \tilde{n})^2}$$

where $\tilde{n} = r\sqrt{\left(\frac{n}{\text{NA}}\right)^2 - 1}$, S_z is the scatter coefficient (11.2) per unit thickness and z is the thickness of the sample.

Experimental subjects. Adult C57BL/6J mice (Jackson laboratory) aged 8–16 weeks old were group housed until surgery and then singly housed on a reverse light cycle. For experiments that did not require surgery, mice were singly housed for at least 5 days before testing. ArcCreER^{T2} mice were group housed until they were 7–9 weeks old when surgeries were performed, after which the mice were singly housed. In the days preceding behavioural experiments, mice were handled regularly to adapt them to the experimenter and the attachment of fibre patch cords. Halorhodopsin experiments began 3–4 weeks after the viral injection to achieve high levels of expression.

Stereotactic surgery. Animals were anaesthetized with ketamine and xylazine (100 mg per kg, 10 mg per kg, respectively, Henry Schein) and placed in a stereotactic frame (Narishige). Small craniotomies were made using standard aseptic technique. Virus was injected with needles pulled from capillary glass (Drummond) at a flow rate of approximately 100 nl min^{-1} by manual pressure injection. Each olfactory bulb received two injections of 1 μl of AAV5-hSyn-eNpHR3.0-eYFP (UNC Viral Core) injected in two locations along the anterior–posterior axis relative to bregma: 5.4 and 6.2 anterior–posterior, 1.1 medial–lateral, -1.4 dorsal–ventral from the bulb surface. Injections into cortical amygdala were also 1 μl in volume at coordinates relative to bregma: -1.7 anterior–posterior, 2.8 medial–lateral, -5.9 dorsal–ventral. Animals exhibited broad expression covering at least 90% of the area of the posterolateral cortical amygdala. For experiments with the ArcCreER^{T2} mice, animals were injected with the Cre-dependent virus: AAV5-ef1 α -DIO-hChR2(H134R)-eYFP (UNC Viral Core). Fibres were implanted into the cortical amygdala at about -1.7 anterior–posterior, 2.8 medial–lateral, -5.7 dorsal–ventral for most experiments; however, for experiments with photoactivation of ChR2 in 2-phenylethanol-responsive neurons, the fibres were implanted slightly more posterior (-2.0 anterior–posterior, 2.9 medial–lateral, -5.7 dorsal–ventral). The fibres were fixed in place using a small amount of clear dental cement (CB Metabond, Parkell) covering the surface of the skull, followed by an outer coating of black Ortho Jet dental acrylic (Lang Dental Manufacturing). Buprenorphine (0.05 mg per kg, Henry Schein) was administered. All injection sites and fibre placements were verified histologically and in rare cases mice were excluded if either were mistargeted. All experiments were conducted according to approved protocols at Columbia University.

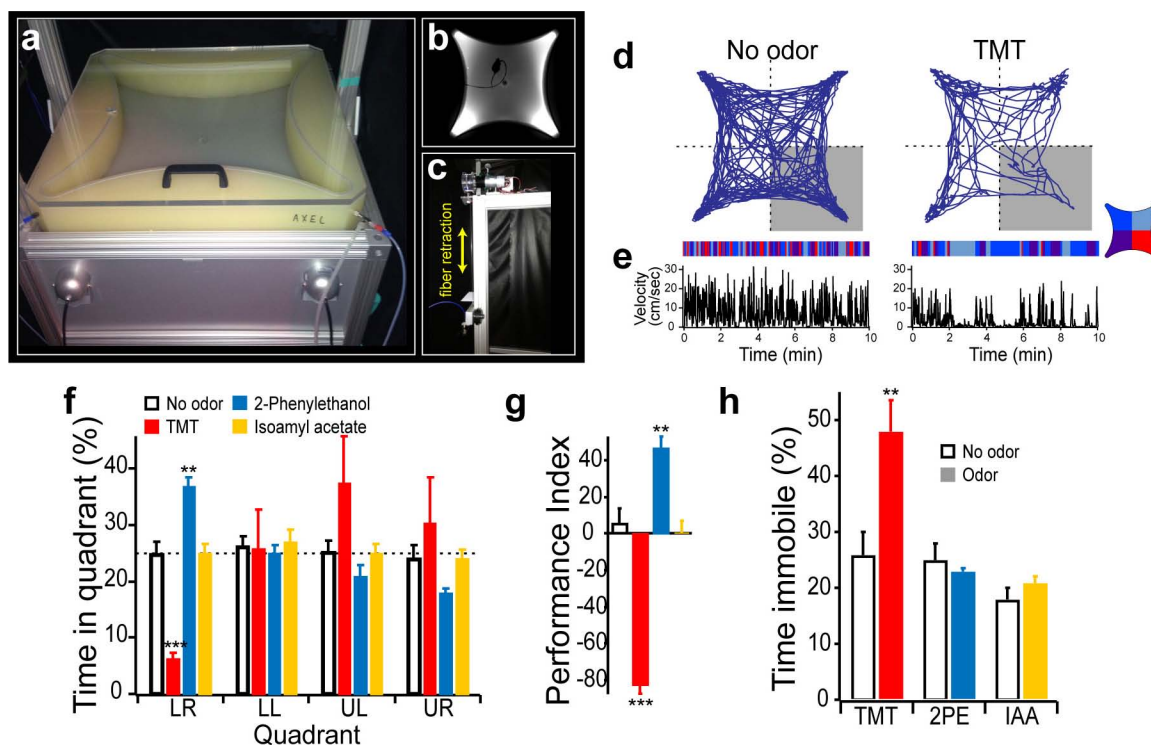
ArcCreER^{T2} mice. A transgenic mouse was generated from a Bac clone containing the entire *arc* gene (see ref. 21). The mice received viral injections at approximately 8 weeks of age. Five days after injection, mice were dark adapted by placing them in a lightproof chamber with fresh bedding for 36–48 h. On day 7 after injection, mice were administered 2 mg of tamoxifen (Sigma T5648). Tamoxifen was prepared as a 10 mg ml^{-1} stock solution dissolved into a 1:10 mixture of ethanol and corn oil (Sigma C8267) and 0.2 ml was administered by intraperitoneal injection. Six hours after tamoxifen, mice were exposed to odour on a cotton swab placed through the roof of their cage twice for 15 min each during a 1 h period. One microlitre of pure odorant was pipetted onto the cotton swab. Mice were maintained in the dark for an additional 48 h after tamoxifen injection before returning to the reverse light cycle. Photoactivation experiments began 3 weeks after tamoxifen and odour exposure.

Histological processing. Mice were administered ketamine and xylazine (100 mg per kg and 10 mg per kg, respectively) and euthanized by transcardial perfusion with 10 ml of PBS, followed by 10 ml of 4% paraformaldehyde in PBS. Brains were extracted and 100 μ m coronal sections were cut on a vibratome. The tissue was labelled with the following antibodies: goat anti-GFP (Abcam ab6673), rabbit anti-arc (Synaptic Systems 156-003), goat anti-c-fos (Santa Cruz sc-52-G), alexa-488 donkey anti-goat (Jackson ImmunoResearch), alexa-568 donkey anti-rabbit (Life Technologies A11057), alexa-647 donkey anti-rabbit (Jackson ImmunoResearch 711-605-152) and alexa-568 donkey anti-goat (Life Technologies A10042). Slices were counterstained with neurotrace 640/660 or 435/455 (Life Technologies N21483 or N21479, respectively). Antibody amplification was not used to visualize eNpHR3.0-eYFP. All images were taken using a Zeiss LSM-710 confocal microscope system.

Cell counting. Confocal images were taken with a $\times 40$ objective and z-stacks of approximately 50 μ m in depth at a resolution of 256 pixels \times 256 pixels. Counting of Chr2-eYFP- and Arc-expressing neurons was performed manually by scrolling through the z-stack with tiled regions of interest. We counted neurons expressing c-fos by creating a projection of the z-stack, and then counting with an automated

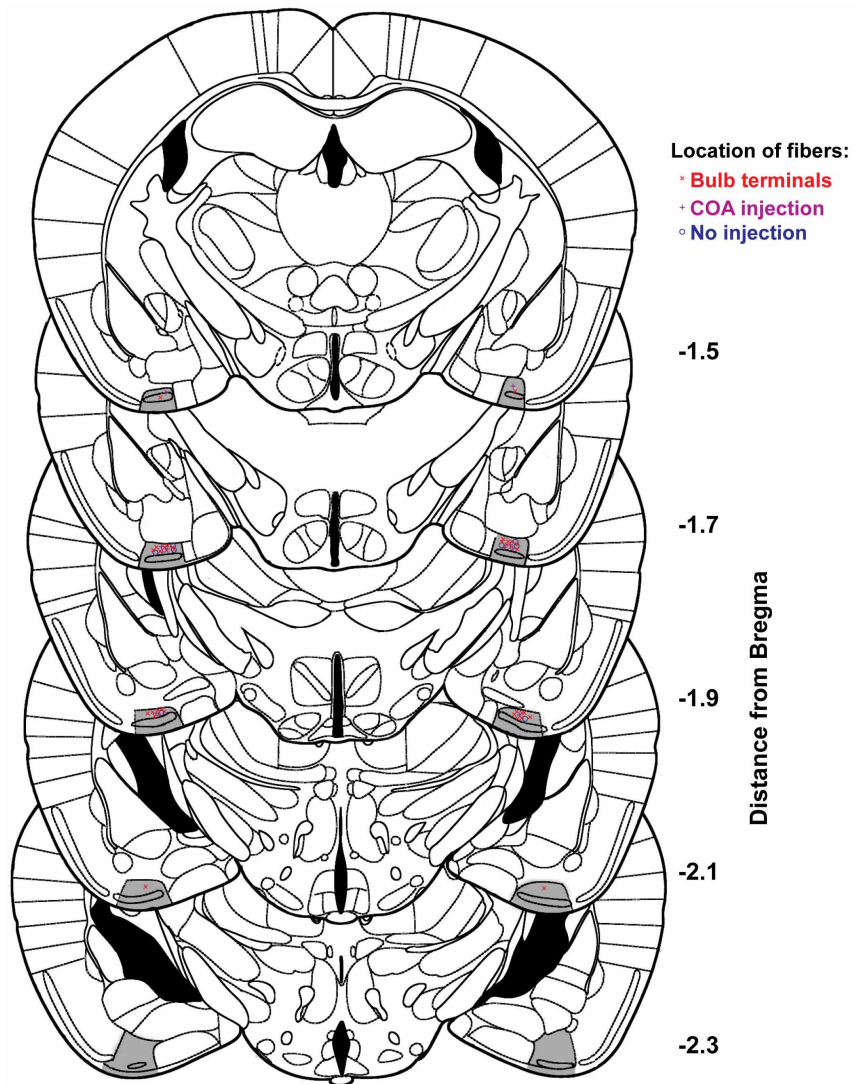
program written in Labview. It was not possible to count the Chr2-eYFP-expressing neurons with the automated program because membrane targeting confounds the detection of cell bodies.

29. Louise, E. M., Vet, J. C. V. L., Heymans, M. & Meelis, E. An airflow olfactometer for measuring olfactory responses of hymenopterous parasitoids and other small insects. *Physiol. Entomol.* **8**, 97–106 (1983).
30. Faucher, C., Forstreuter, M., Hilker, M. & de Bruyne, M. Behavioral responses of *Drosophila* to biogenic levels of carbon dioxide depend on life-stage, sex and olfactory context. *J. Exp. Biol.* **209**, 2739–2748 (2006).
31. Semmelhack, J. L. & Wang, J. W. Select *Drosophila* glomeruli mediate innate olfactory attraction and aversion. *Nature* **459**, 218–223 (2009).
32. Vo-Dinh, T. *Biomedical Photonics Handbook* Ch. 2, 37 (CRC, 2003).
33. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in neural systems. *Neuron* **71**, 9–34 (2011).
34. Aravanis, A. M. *et al.* An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J. Neural Eng.* **4**, S143–S156 (2007).



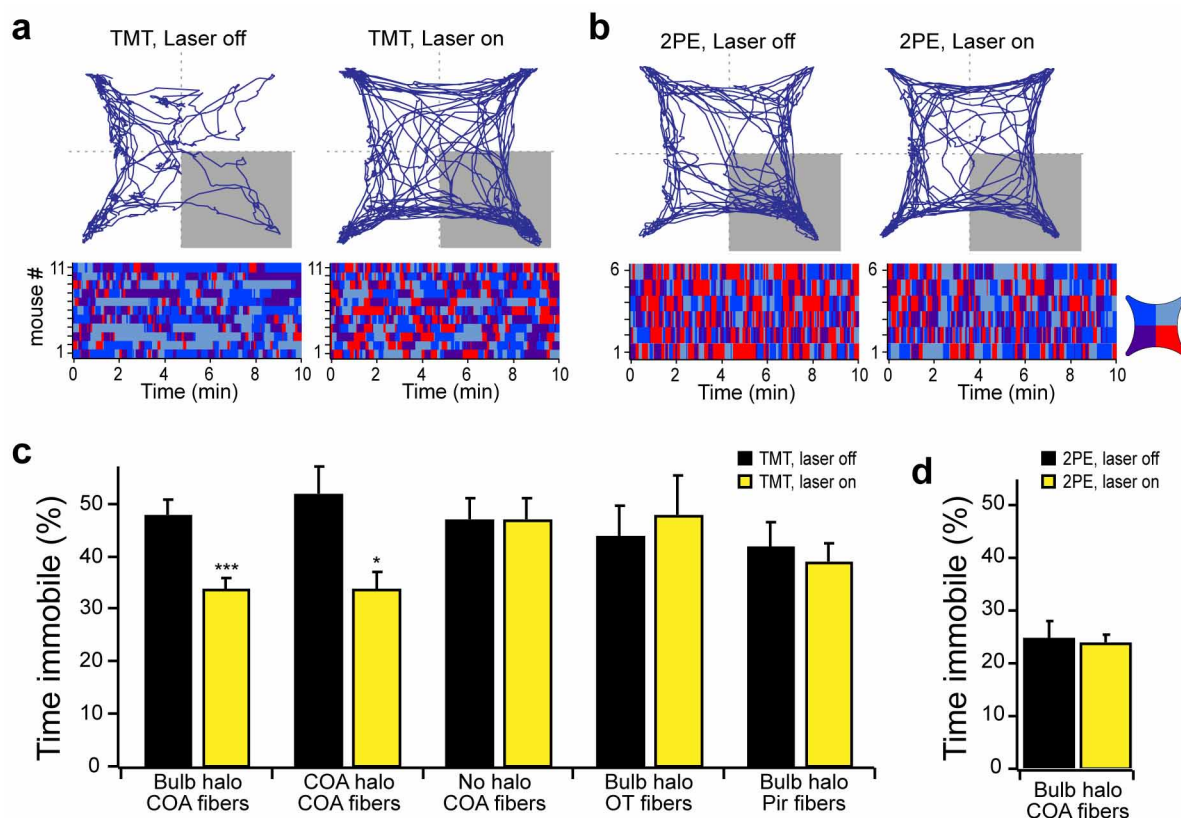
Extended Data Figure 1 | Quantification of the four-quadrant behaviour assay. **a–c**, Images of the four-field behaviour chamber. **b**, An image, taken from the camera that tracked the position of the animal, showing a mouse tethered to optic fibres that entered the chamber through a port on the left side. **c**, A motorized fibre retraction system adjusted the length of the fibre as the mouse traversed the arena. The trajectory of a representative mouse is plotted for a 10 minute period in the absence of odour (left), or after the addition of odour to the lower right quadrant (right). The raster plots below the trajectory graphs represent quadrant occupancy over time. **e**, The velocity over time in the absence of odour (left) or in the presence of TMT (right) reveals bouts of

inactivity associated with freezing behaviour in the presence of TMT. **f**, The average amount of time spent in each quadrant either in the absence of odour, or the presence of TMT, 2-phenylethanol or isoamyl acetate. **g**, This quantification is reduced when plotted as the performance index ($n = 5$). **h**, Pauses in locomotor activity are quantified as the percentage time immobile in the presence and absence of TMT, 2-phenylethanol or isoamyl acetate. Immobility is defined as velocity less than 1 cm s^{-1} for at least 1 s. **f–h**, ** $P < 0.01$, *** $P < 0.001$, paired t -test comparing performance index with and without odour for each odour group; error bars, s.e.m.



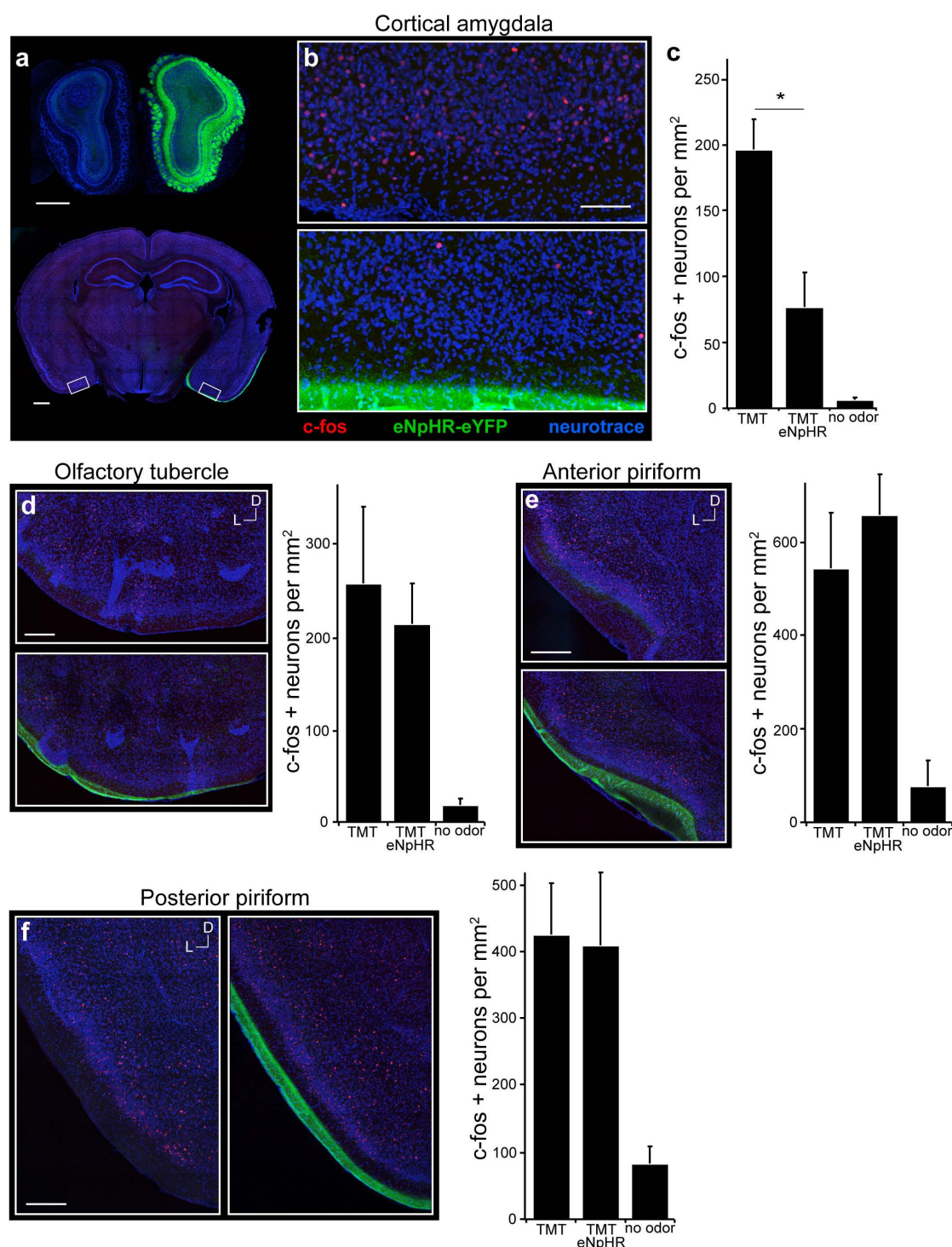
Extended Data Figure 2 | Location of optical fibres implanted in cortical amygdala for photoactivation of halorhodopsin. Schematics show coronal sections throughout most of the region containing cortical amygdala. The

posterolateral cortical amygdala is highlighted in grey and the location of bilaterally implanted fibres is indicated.



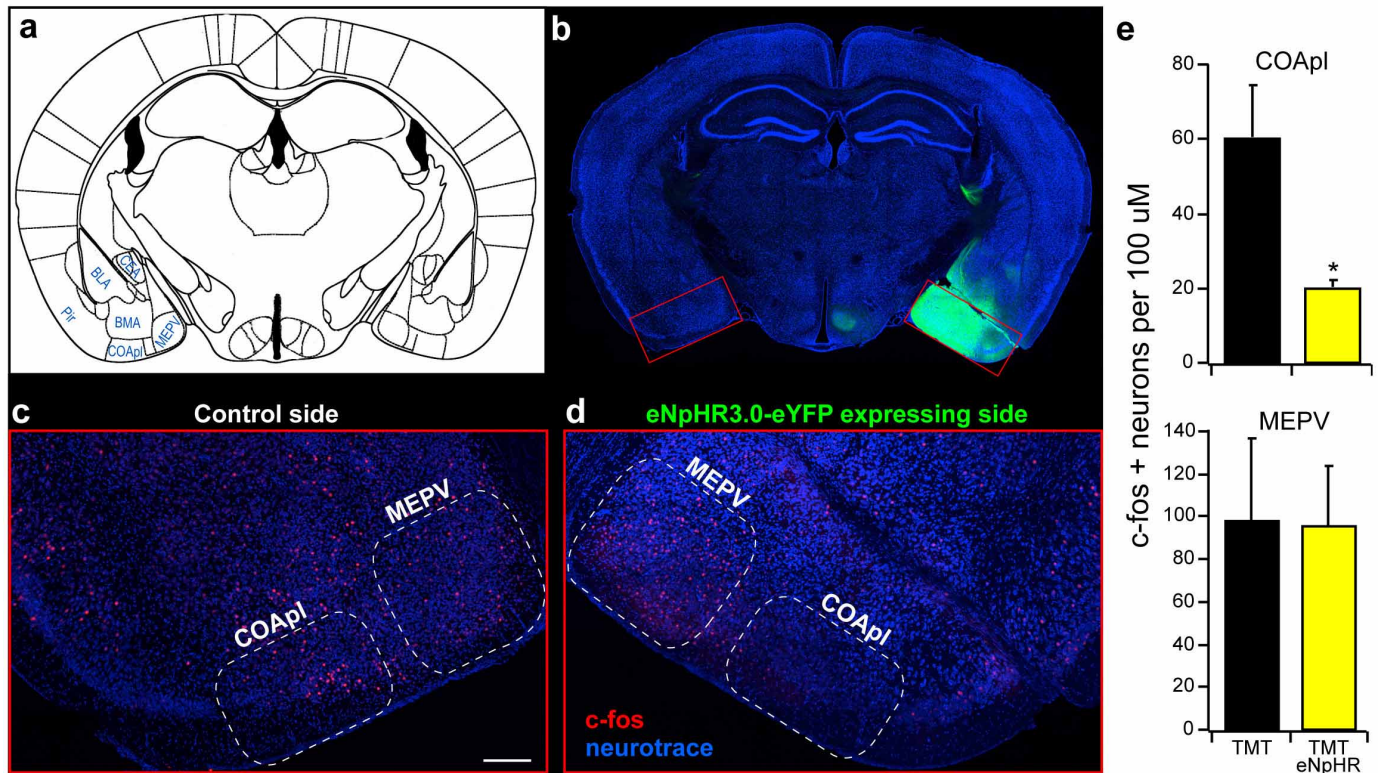
Extended Data Figure 3 | Locomotor activity of mice with optical suppression in cortical amygdala. **a, b,** Mice with halorhodopsin in the olfactory bulb and optical fibres in cortical amygdala were optically coupled to a yellow laser and tested in the behavioural assay for the response to TMT (**a**) or 2-phenylethanol (**b**) with and without laser stimulation. The position of a representative mouse during a 10 minute period in the presence of TMT (**a**) or 2-phenylethanol (**b**) either in the absence (left) or presence (right) of photoactivation during the 10 minute behavioural testing is shown. Raster plots show quadrant occupancy over time for each animal (**a**, $n = 11$; **b**, $n = 6$). **c, d,** The percentage time immobile in the absence and presence of photoactivation. Immobility is defined as velocity less than 1 cm s^{-1} for at least

1 s. **c,** Response to TMT in mice receiving photostimulation of halorhodopsin in different experimental animals. Bulb halo and COA halo describe mice with halorhodopsin expression in the olfactory bulb and cortical amygdala, respectively. Optical fibres were placed above cortical amygdala (COA, $n = 11$), olfactory tubercle (OT, $n = 7$) or in piriform cortex (Pir, $n = 8$) as denoted below the site of injection. Control animals received no viral injection, and fibres implanted into cortical amygdala ($n = 4$). **d,** The percentage immobility for mice exposed to 2-phenylethanol in the absence and presence of photoactivation of bulbar axons in cortical amygdala ($n = 6$). **c, d,** $*P < 0.05$, $***P < 0.001$, paired t -test comparing with and without laser; error bars, s.e.m.



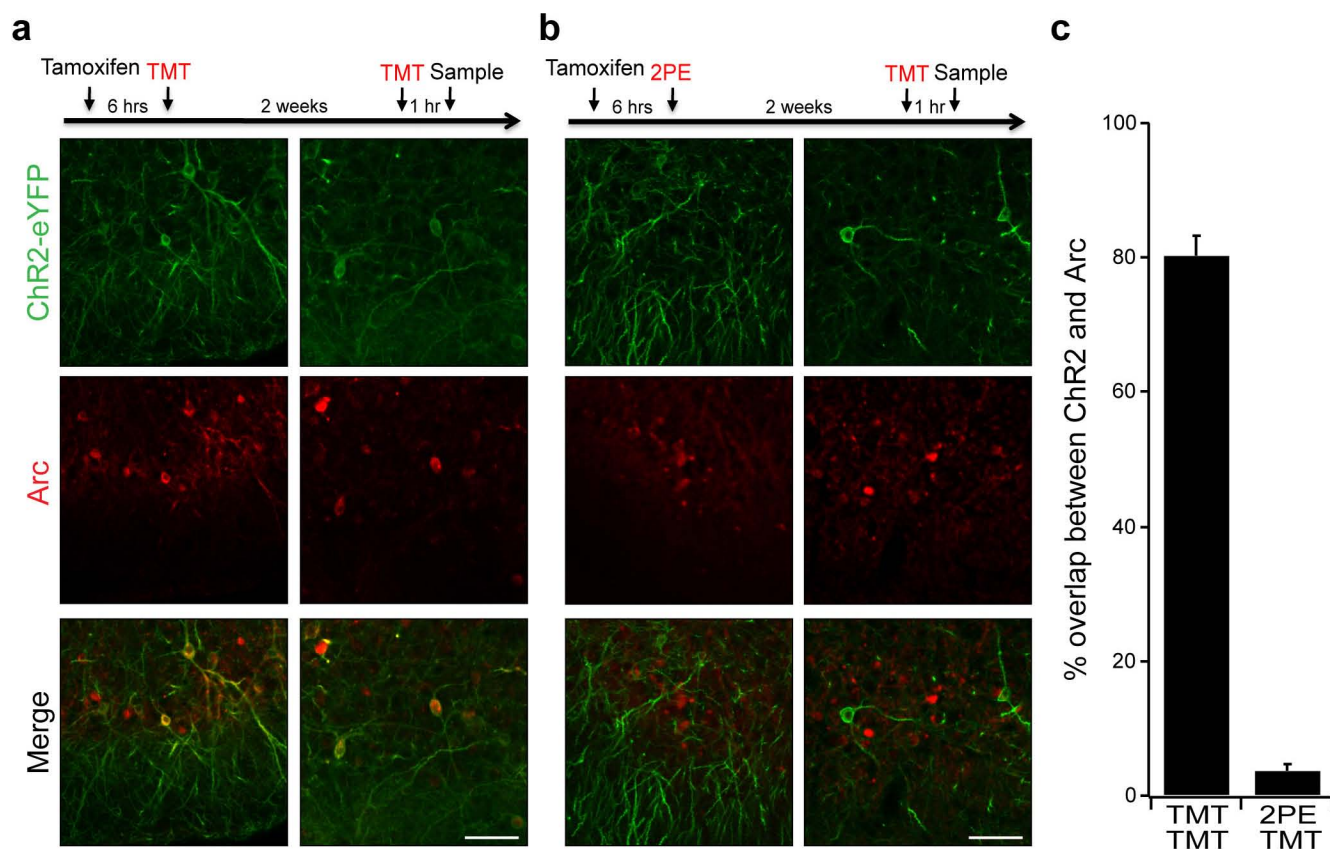
Extended Data Figure 4 | Optical suppression of bulbar input to cortical amygdala selectively reduces odour-evoked activity in this region. Mice were unilaterally injected with AAV5-eNpHR3.0-eYFP into the olfactory bulb and exposed to TMT while photoactivating the mitral-tufted cell axon terminals in cortical amygdala. **a**, Coronal section of the olfactory bulbs (top) and of the brain region with cortical amygdala (bottom) reveal halorhodopsin expression in the bulb and the lateral olfactory tract. **b**, Magnified and cropped image showing the cortical amygdala of both sides of the posterolateral cortical amygdala. The region in the bottom image received photoactivation of halorhodopsin in bulbar axon terminals during odour exposure, whereas the contralateral side (top) did not. Scale bars, 200 μm (**a**) and 100 μm (**b**). **c**, The number of neurons expressing c-fos was counted for each side of the brain

($n = 4$) as well as in control animals that did not receive any stimuli ($n = 2$) across multiple regions of cortical amygdala. Numbers are normalized by area (square millimetres) of 100 μm thick sections for comparison between brain areas. * $P < 0.05$, paired t -test comparing with and without laser; error bars, s.e.m. **d–f**, Coronal sections showing the olfactory tubercle (**d**), anterior piriform (**e**) and posterior piriform (**f**); the region in the bottom images received photoactivation of bulbar axon terminals in the cortical amygdala. Scale bars, 200 μm . Bar graphs show the number of neurons expressing c-fos counted for each side of the brain ($n = 4$) as well as in control animals that did not receive any stimuli ($n = 2$). Error bars, s.e.m. **a–f**, Images are taken from sections at the following anterior–posterior distances from bregma: -1.7 (**b**), 1.2 (**d**, **e**), -1.6 (**f**).



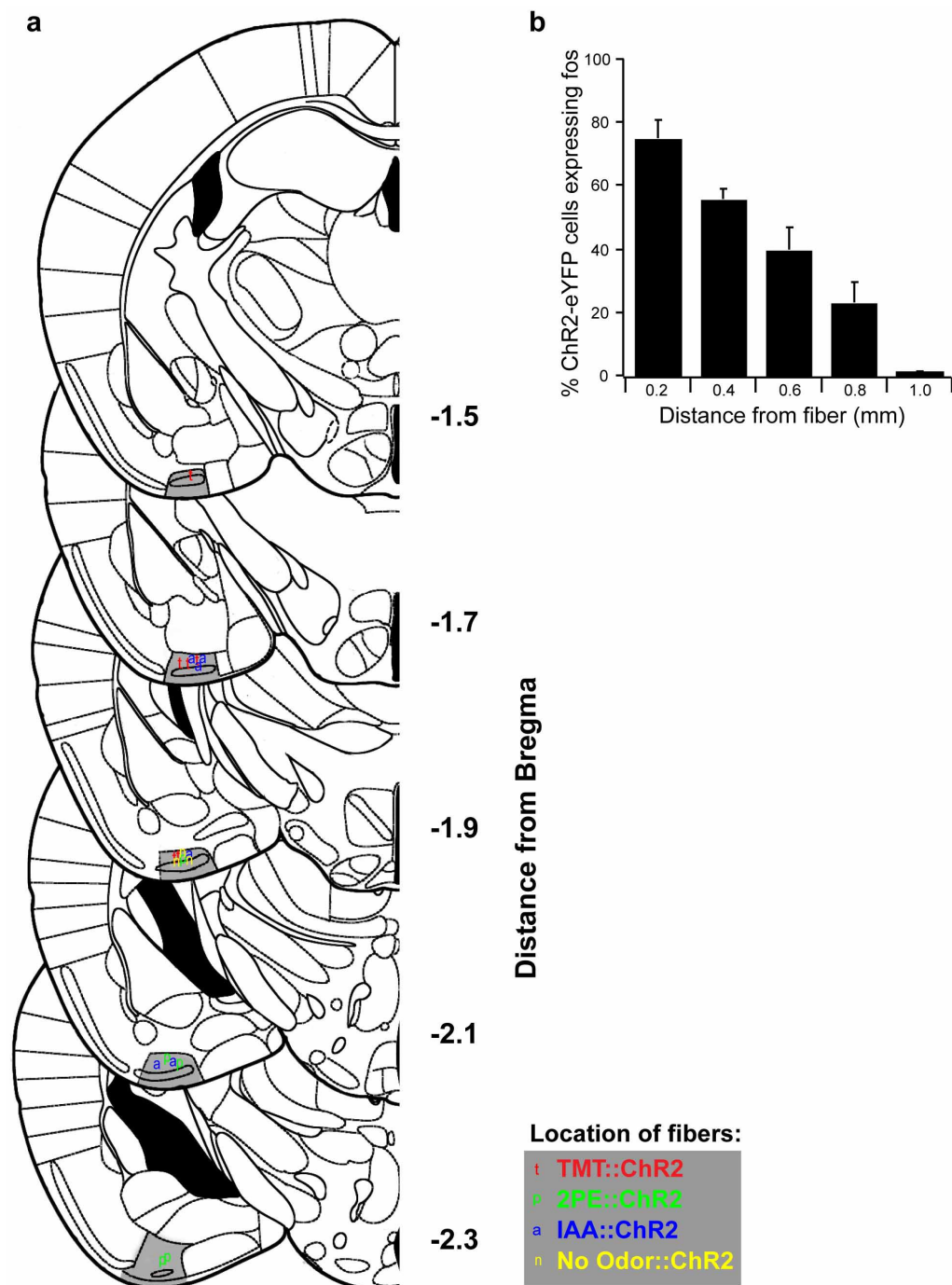
Extended Data Figure 5 | Halorhodopsin expression of neurons within cortical amygdala. Injection of AAV into cortical amygdala leads to broad expression of halorhodopsin within cortical amygdala and neighbouring areas; however, optical silencing is restricted to the cortical amygdala. **a**, Schematic of a coronal section showing cortical amygdala (COApl) in relation to other ventral brain regions. **b–d**, Mice were unilaterally injected with AAV5-eNpHR3.0-eYFP into the cortical amygdala and exposed to TMT while photoactivating halorhodopsin in cortical amygdala neurons. **b**, Coronal section reveals broad expression of halorhodopsin in the cortical amygdala

region. This expression is broad, covering at least 90% of posterolateral cortical amygdala throughout the anterior–posterior axis. **c, d**, Magnified and cropped image showing c-fos expression in the cortical amygdala. The right side (**d**) received photoactivation during odour exposure, whereas the contralateral side (**c**) did not. Scale bar, 200 μm. **e**, The number of neurons expressing c-fos was counted for each side of the brain ($n = 3$) across multiple regions of the cortical amygdala (top) as well as the medial amygdala (MEPV, bottom). The mean cell number for each animal is shown in the bar graph. * $P < 0.05$, paired t -test comparing with and without laser stimulation; error bars, s.e.m.



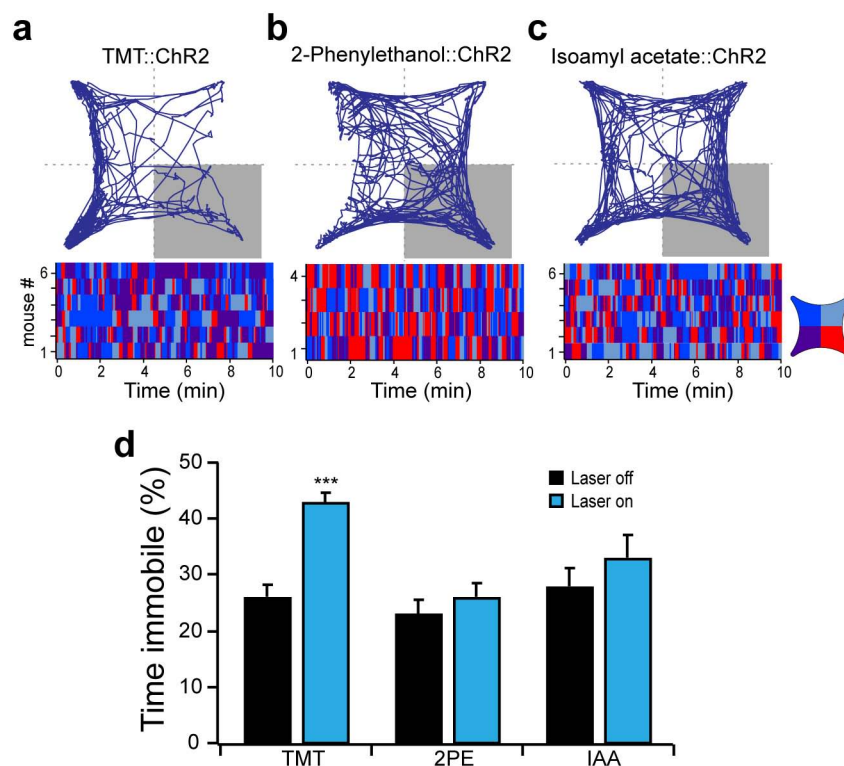
Extended Data Figure 6 | The arc promoter can be used to drive channelrhodopsin expression faithfully in odour-specific neurons. Arc-Cre-ER^{T2} mice were administered tamoxifen and exposed to odour to induce ChR2-eYFP expression, and subsequently exposed to either the same odour or a different odour and then sampled for endogenous arc expression. **a, b**, The timeline for odour exposure is indicated at the top, and images from two different regions of cortical amygdala are shown from representative animals

for each experiment. **a**, TMT exposure induced ChR2-eYFP expression in the cortical amygdala and re-exposure to the same odour-induced arc expression detected by immunocytochemistry. **b**, 2-Phenylethanol induced expression of ChR2-eYFP followed by re-exposure to TMT. Scale bar, 100 μ m. **c**, The number of channelrhodopsin-expressing neurons that also expressed the endogenous arc protein. Neurons were counted across randomly chosen sections throughout the cortical amygdala ($n = 4$).



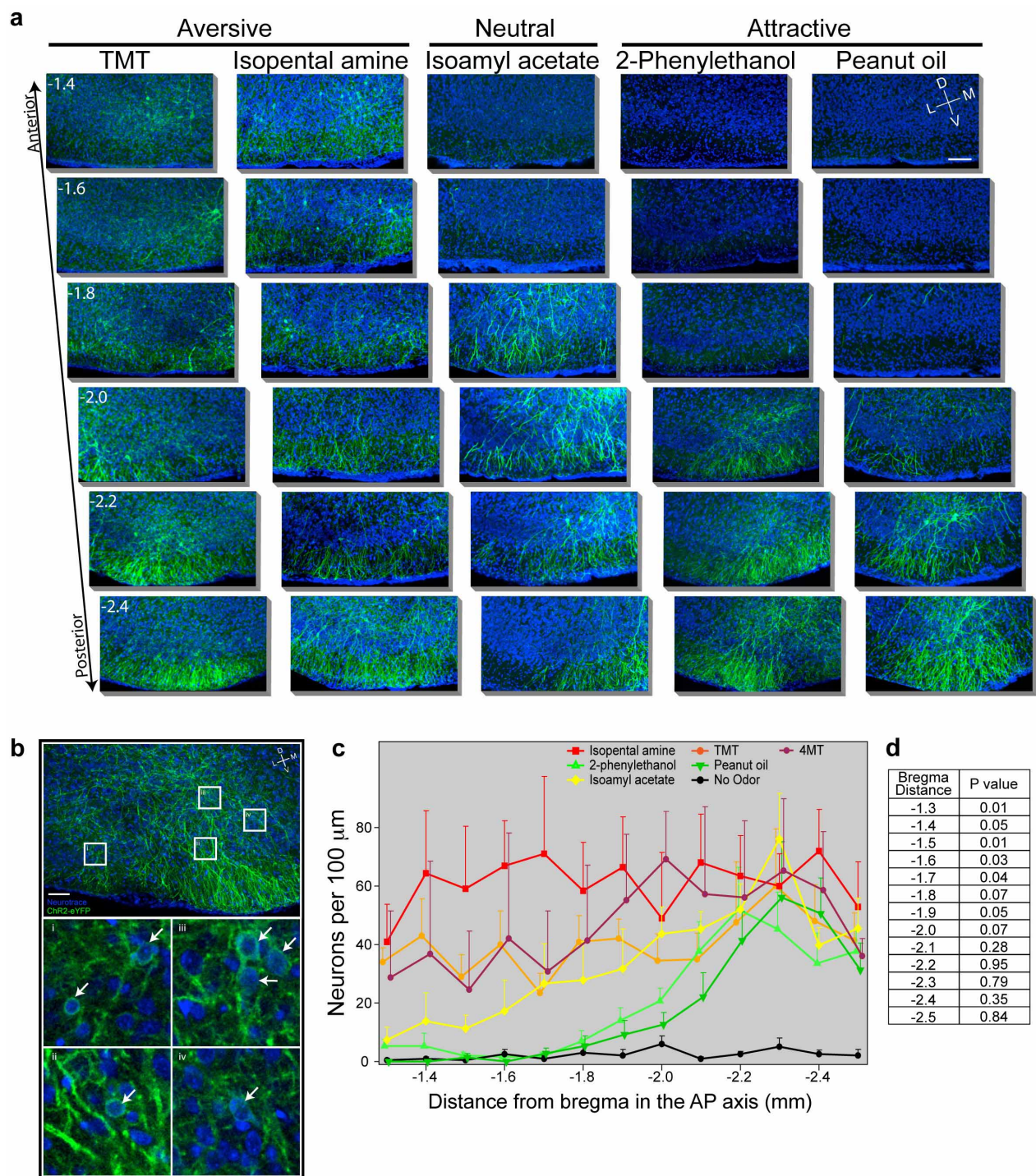
Extended Data Figure 7 | Location of optical fibres implanted in cortical amygdala for photoactivation of odour-responsive neurons. **a**, Schematics show unilateral coronal sections throughout most of the region containing cortical amygdala. The posterolateral cortical amygdala is highlighted in grey and the location of unilaterally implanted fibres is indicated. Fibres were not preferentially targeted to one side of the brain, but the fibre positions are collapsed onto unilateral schematics. **b**, The extent of light-induced activation of channelrhodopsin-expressing neurons as a function of distance from the

fibre tip using c-fos expression. ArcCreER^{T2} mice were injected with AAV5-eYFP- α -DIO-ChR2-eYFP into cortical amygdala, administered tamoxifen and exposed to TMT to induce channelrhodopsin expression. Three weeks later, the cortical amygdala was photoactivated for 10 min with cycles of 30 s of pulsed light (10 Hz, 50% duty cycle) and 30 s off. Mice were sampled for c-fos immunoreactivity as a function of distance from the fibre tip in 200 μ m bins ($n = 3$). $P < 0.001$, one-way ANOVA; error bars, s.e.m.



Extended Data Figure 8 | Locomotor activity of mice during activation of odour-responsive neurons within cortical amygdala. Mice with odour-driven channelrhodopsin expression were tested in the open field assay where they received pulsed photoactivation upon entrance into the lower right quadrant. **a–c**, The trajectory graphs (top) show the position of representative animals with ChR2–eYFP in neurons activated by TMT (**a**), 2-phenylethanol

(**b**) or isoamyl acetate (**c**). The raster plots (bottom) show quadrant occupancy over time. **d**, The percentage time immobile in the absence and presence of photoactivation. Immobility is defined as velocity less than 1 cm s^{-1} for at least 1 s. **a–c**, TMT ($n = 6$), 2-phenylethanol ($n = 4$) and isoamyl acetate ($n = 6$); *** $P < 0.001$, paired t -test comparing with and without laser; error bars, s.e.m.



Extended Data Figure 9 | The spatial distribution of neurons responsive to different odours within cortical amygdala. **a**, ArcCreER^{T2} mice were administered tamoxifen and exposed to one of five different odours and then consecutive, serial, coronal sections were collected throughout the cortical amygdala region. Images reveal odour-driven ChR2-eYFP expression in serial sections across the cortical amygdala of representative mice for each odour. Images of the ventral brain region are magnified and cropped to show only the posterolateral cortical amygdala, and sections are displayed at 200 μm intervals across 1.2 mm of the anterior-posterior axis. Scale bar, 100 μm . **b**, Representative images showing the expression of odour-induced ChR2-eYFP with magnified images that reveal identifiable cell bodies for counting.

The top image shows a z-projection of 40 μm through cortical amygdala (left), and the bottom images (i-iv) show magnified single z-plane images of small areas revealing neuronal cell bodies, indicated by the white arrows. Scale bar, 60 μm . **c**, The average number of neurons counted per 100 μm coronal section throughout the posterolateral cortical amygdala for different odours. Error bars, s.e.m. **d**, A one-way ANOVA was performed for each point along the anterior-posterior axis, comparing the means between different odours (excluding the no-odour condition). The no-odour condition was compared with each odour at -2.3 from bregma using an unpaired *t*-test; $P < 0.001$ for each odour.

A three-dimensional human neural cell culture model of Alzheimer's disease

Se Hoon Choi^{1*}, Young Hye Kim^{1,2*}, Matthias Hebesch^{1,3}, Christopher Sliwinski¹, Seungkyu Lee⁴, Carla D'Avanzo¹, Hechao Chen¹, Basavaraj Hooli¹, Caroline Asselin¹, Julien Muffat⁵, Justin B. Klee¹, Can Zhang¹, Brian J. Wainger⁴, Michael Peitz³, Dora M. Kovacs¹, Clifford J. Woolf⁴, Steven L. Wagner⁶, Rudolph E. Tanzi¹ & Doo Yeon Kim¹

Alzheimer's disease is the most common form of dementia, characterized by two pathological hallmarks: amyloid- β plaques and neurofibrillary tangles¹. The amyloid hypothesis of Alzheimer's disease posits that the excessive accumulation of amyloid- β peptide leads to neurofibrillary tangles composed of aggregated hyperphosphorylated tau^{2,3}. However, to date, no single disease model has serially linked these two pathological events using human neuronal cells. Mouse models with familial Alzheimer's disease (FAD) mutations exhibit amyloid- β -induced synaptic and memory deficits but they do not fully recapitulate other key pathological events of Alzheimer's disease, including distinct neurofibrillary tangle pathology^{4,5}. Human neurons derived from Alzheimer's disease patients have shown elevated levels of toxic amyloid- β species and phosphorylated tau but did not demonstrate amyloid- β plaques or neurofibrillary tangles^{6–11}. Here we report that FAD mutations in β -amyloid precursor protein and presenilin 1 are able to induce robust extracellular deposition of amyloid- β , including amyloid- β plaques, in a human neural stem-cell-derived three-dimensional (3D) culture system. More importantly, the 3D-differentiated neuronal cells expressing FAD mutations exhibited high levels of detergent-resistant, silver-positive aggregates of phosphorylated tau in the soma and neurites, as well as filamentous tau, as detected by immunoelectron microscopy. Inhibition of amyloid- β generation with β - or γ -secretase inhibitors not only decreased amyloid- β pathology, but also attenuated tauopathy. We also found that glycogen synthase kinase 3 (GSK3) regulated amyloid- β -mediated tau phosphorylation. We have successfully recapitulated amyloid- β and tau pathology in a single 3D human neural cell culture system. Our unique strategy for recapitulating Alzheimer's disease pathology in a 3D neural cell culture model should also serve to facilitate the development of more precise human neural cell models of other neurodegenerative disorders.

To develop human neural progenitor cells (hNPCs) that produce high levels of toxic amyloid- β species, we overexpressed human β -amyloid precursor protein (APP) or APP and presenilin 1 (PSEN1), containing FAD mutations. We first generated polycistronic lentiviral constructs designed to express human APP with both K670N/M671L (Swedish) and V717I (London) FAD mutations (APPSL) or APPSL and PSEN1 with Δ E9 FAD mutation (PSEN1(Δ E9)) (Fig. 1a). These FAD lentiviral constructs were transfected into ReNcell VM human neural stem (ReN) cells (Millipore). The transfected cells with GFP (ReN-G), mCherry (ReN-m), APPSL-GFP (ReN-GA), APPSL-GFP/PSEN1(Δ E9)-mCherry (ReN-mGAP), APPSL-PSEN1(Δ E9)-mCherry (ReN-mAP) or GFP/APPSL-PSEN1(Δ E9)-mCherry (HReN-mGAP) were enriched on the basis of GFP and/or mCherry signals using fluorescence-activated cell sorting (FACS) (Extended Data Fig. 1a–c, f). Western blot analysis revealed high expression of PSEN1(Δ E9), APP and APP carboxy-terminal fragments in ReN cells with FAD mutations (FAD ReN cells, Extended Data Fig. 1e).

As previously reported^{12–14}, most ReN cells differentiated into neuronal and glial cells within 3 weeks (Extended Data Figs 1d and 2a). Immunofluorescence staining confirmed punctate localization of VGLUT1 (vesicular glutamate transporter 1), a presynaptic marker, along with dendritic processes (Extended Data Fig. 2b). Quantitative real-time PCR (qPCR) array analysis showed robust increases of neuronal and glial marker genes (Extended Data Fig. 2c). Reverse-transcription-PCR (RT-PCR) analysis also showed a dramatic increase of 4-repeat adult tau isoforms following differentiation (Extended Data Fig. 2d). The differentiated ReN cells

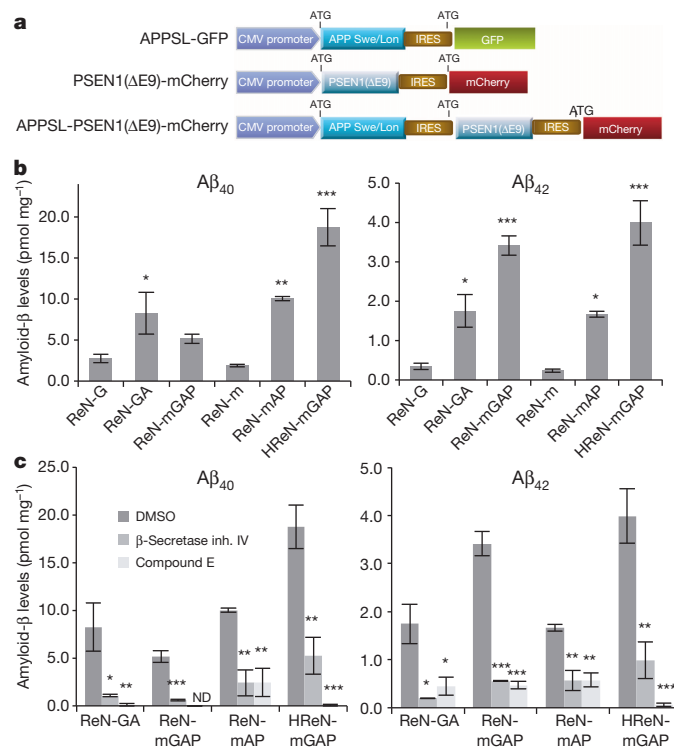


Figure 1 | Generation of hNPCs with multiple FAD mutations. **a**, Diagrams showing lentiviral internal ribosome entry sites (IRES) constructs. CMV, cytomegalovirus. **b**, Increased A β ₄₀ and A β ₄₂ levels in 6-week differentiated FAD ReN cells. Amyloid- β levels in conditioned media were normalized to total protein levels. * P < 0.05; ** P < 0.01; *** P < 0.001; ANOVA followed by a post hoc Dunnett's test; n = 3 per each sample. **c**, Amyloid- β levels are dramatically decreased in FAD ReN cells after treatment with 1 μ M β -secretase inhibitor IV or 3.7 nM compound E. Mean \pm s.e.m.; * P < 0.05; ** P < 0.01; *** P < 0.001; ANOVA followed by a post hoc Dunnett's test; n = 3 per each sample; ND, not detected.

¹Genetics and Aging Research Unit, MassGeneral Institute for Neurodegenerative Disease, Massachusetts General Hospital, Harvard Medical School, Charlestown, Massachusetts 02129, USA. ²Division of Mass Spectrometry Research, Korea Basic Science Institute, Cheongju-si, Chungbuk 363-883, South Korea. ³Institute of Reconstructive Neurobiology, Life and Brain Center, University of Bonn and Hertie Foundation, 53127 Bonn, Germany. ⁴FM Kirby Neurobiology Center, Boston Children's Hospital and Harvard Stem Cell Institute, Boston, Massachusetts 02115, USA. ⁵The Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. ⁶Department of Neurosciences, University of California, San Diego, La Jolla, California 92093, USA.

*These authors contributed equally to this work.

exhibit voltage-gated potassium and sodium currents after 29 days of differentiation (Extended Data Fig. 2e–h), as previously reported^{12,13}.

We measured levels of the 40-amino-acid and 42-amino-acid amyloid- β isoforms ($A\beta_{40}$ and $A\beta_{42}$) in conditioned media after 6-week differentiation. FAD ReN cells revealed dramatic increases in $A\beta_{40}$ (~9-fold) and $A\beta_{42}$ (~17-fold) levels as compared to the control ReN cells (Fig. 1b and Extended Data Fig. 2i). The $A\beta_{42}:A\beta_{40}$ ratio was also increased (~5-fold) in ReN cells expressing PSEN1(Δ E9) (Fig. 1b). Treatment with β - or γ -secretase inhibitors markedly decreased amyloid- β levels (Fig. 1c) with no appreciable toxicity (data not shown). We confirmed that ReN

cells carry the APOE $\epsilon 3/\epsilon 3$ genotype, not the $\epsilon 4$ allele associated with Alzheimer's disease (Extended Data Fig. 2j).

In conventional 2D cultures, secreted amyloid- β diffuses into a large volume of media. We hypothesized that a 3D culture would accelerate amyloid- β deposition by limiting diffusion of amyloid- β , allowing for aggregation^{15–17}. We chose BD Matrigel (BD Biosciences) as a 3D support matrix since it contains high levels of brain extracellular matrix proteins. For immunofluorescence and biochemical analyses, we set up thin- (~100–300 μ m) and thick-layer (~4 mm) 3D culture models (Figs 2a and 3a). 3D-differentiated ReN cells displayed extensive processes after

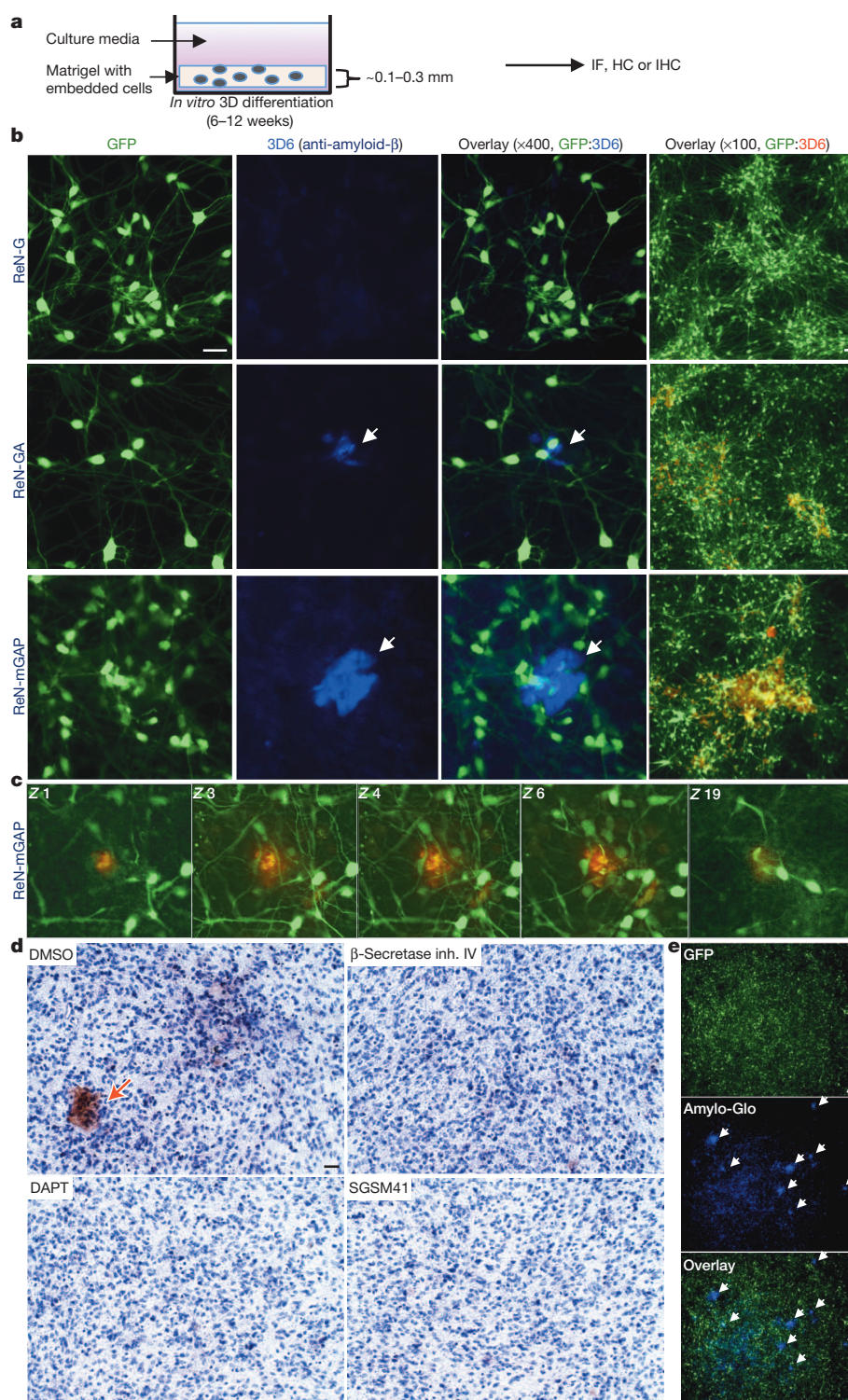


Figure 2 | Robust increases of extracellular amyloid- β deposits in 3D-differentiated hNPCs with FAD mutations. **a**, Thin-layer 3D culture protocol. HC, histochemistry; IF, immunofluorescence; IHC, immunohistochemistry. **b**, Amyloid- β deposits in 6-week differentiated control and FAD ReN cells in 3D Matrigel (green, GFP; blue, 3D6; scale bar, 25 μ m; arrowheads, extracellular amyloid- β deposits; right-most panels, 3D6 staining was pseudo-coloured to red). **c**, Select confocal Z-stack images of 3D6-positive amyloid- β deposits. Z-sections with an interval of 2 μ m were captured and sections 1, 3, 4, 6 and 19 are shown (green, GFP; red, 3D6). **d**, IHC of amyloid- β deposits in ReN-mGAP cells. 3D-differentiated cells were treated with 1 μ M β -secretase inhibitor IV, 500 nM DAPT, 500 nM SGSM41 or dimethyl sulphoxide (DMSO). Brown, 3,3'-diaminobenzidine (DAB, BA27); blue, haematoxylin; scale bar, 25 μ m; arrowhead, large amyloid- β deposits. **e**, Detection of amyloid plaques in ReN-mGAP cells with Amylo-Glo (green, GFP; blue, Amylo-Glo; arrows, Amylo-Glo-positive aggregates).

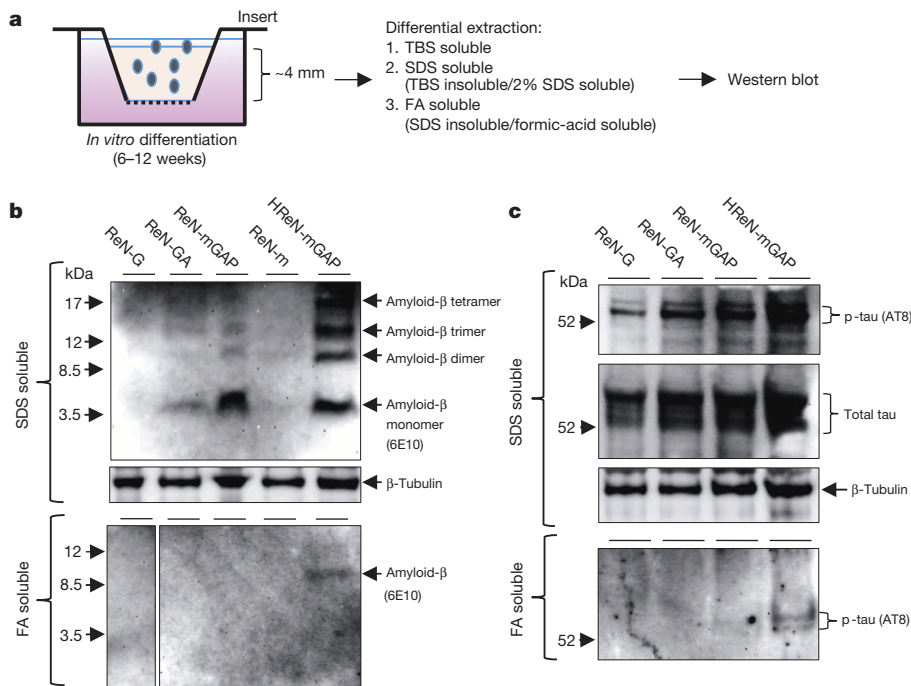


Figure 3 | Elevation of amyloid- β and p-tau levels in TBS-insoluble fractions of 3D-differentiated FAD hNPCs. a, A diagram showing a thick-layer 3D culture and detergent extraction protocols. **b**, Western blot of amyloid- β aggregates in 3D-differentiated ReN cells. 6E10-antibody-detected amyloid- β monomers, dimers, trimers and tetramers in SDS-soluble (upper panel) and formic-acid-soluble fractions (lower panel) from the control (ReN-G and -m) and the FAD ReN cells (ReN-GA, ReN-mGAP and HReN-mGAP) after 6 weeks of differentiation. **c**, Western blot of total and p-tau levels in SDS-soluble and formic-acid-soluble fractions.

2–6 weeks of differentiation (Supplementary Videos 1, 2 and 3). To demonstrate ReN-cell differentiation in thick-layer 3D cultures, paraffin-sectioned thick-layer 3D cultures were analysed by immunofluorescence with various neuronal markers (Extended Data Fig. 3). qPCR array data indicated that 3D cultures promote more neuronal and glial differentiation than 2D cultures (Extended Data Fig. 2c). We also observed a dramatic increase in the levels of 4-repeat adult tau isoforms in 3D cultures versus 2D (Extended Data Fig. 2d). These data demonstrate that 3D culture conditions not only promote neuronal maturation, but also increase adult tau isoform levels, which are essential for reconstituting tauopathy¹⁸.

To detect amyloid- β deposits, 3D-differentiated ReN cells were stained with the anti-amyloid- β antibody 3D6 (Fig. 2b). Notably, confocal microscopy showed robust increases in extracellular amyloid- β deposits in FAD ReN cells (Fig. 2b). Confocal Z-sectioning exhibited 3D extracellular amyloid- β aggregates with diameters of 10–50 μ m (Fig. 2c and Supplementary Video 4). Immunohistochemistry (IHC) with the BA27 anti-A β_{40} antibody reconfirmed robust increases in amyloid- β aggregates and diffuse amyloid- β staining in FAD ReN cells (Fig. 2d and Extended Data Fig. 4a–c). Treatment with β - or γ -secretase inhibitors dramatically decreased amyloid- β deposits (Fig. 2d and Extended Data Fig. 4a). Accumulation of insoluble amyloid- β aggregates was also confirmed by Amylo-Glo, a fluorescent amyloid-specific dye (Biosensis; Fig. 2e and Extended Data Fig. 4d) and Congo red staining (data not shown). Treatment with the γ -secretase modulator SGSM41 (ref. 19) (Extended Data Table 1) markedly decreased the number of dense amyloid- β aggregates in 3D-differentiated FAD ReN cells, possibly by selectively decreasing the aggregation-prone A β_{42} species (Fig. 2d and Extended Data Fig. 4a). For biochemical analysis, 3D-differentiated ReN cells were homogenized in Tris-buffered saline (TBS) and underwent a serial extraction protocol with 2% sodium dodecyl sulphate (SDS) and formic acid (Fig. 3a–c). Western blot analysis with the anti-amyloid- β antibody 6E10 revealed increased amyloid- β levels in SDS-soluble fractions of FAD ReN cells (Fig. 3b). In addition to a ~4-kDa amyloid- β monomer band, we observed amyloid- β dimer, trimer and tetramer bands in FAD ReN cells with high A β_{42} levels (Fig. 3b). Amyloid- β dimers were detected in formic-acid-soluble fractions, especially in HReN-mGAP cells, which generated the highest levels of A β_{40} and A β_{42} (Fig. 3b).

In Alzheimer's disease, tau protein is hyperphosphorylated and abnormally accumulates in axons, dendrites and cell bodies²⁰. To analyse

phosphorylated tau (p-tau) levels in FAD ReN cells, we performed western blot analysis with antibodies against p-tau (phosphorylated at Ser 199/Ser 202/Thr 205, AT8), and showed that p-tau levels were strongly increased in SDS-soluble fractions (Fig. 3c). Species of p-tau were also detected in formic-acid-soluble fractions in HReN-mGAP cells, which showed the highest accumulation of insoluble amyloid- β (Fig. 3b, c). Moderate increases of total tau levels were detected in FAD ReN cells, particularly in HReN-mGAP cells (Fig. 3c), which might be explained by robust increases in aggregated tau fractions. Notably, the γ -secretase inhibitor DAPT not only blocked amyloid- β generation but also inhibited p-tau levels in HReN-mGAP cells (Extended Data Fig. 5a, b). For further exploration of p-tau accumulation at the cellular level, we performed IHC staining with two different p-tau antibodies, AT8 and PHF1 (p-tau phosphorylated at Ser 396/Ser 404). We found dramatic increases in p-tau levels in a small portion of FAD ReN cells (Extended Data Fig. 5c–g). These cells displayed unusual morphologies, including beaded processes that are similar to the dystrophic neurites found in the brains of Alzheimer's disease patients (Extended Data Fig. 5c–e). Additionally, areas with highly dense p-tau-positive processes could only be detected in FAD ReN cells (Extended Data Fig. 5e, f). Treatment with either β - or γ -secretase inhibitors decreased p-tau-positive cell numbers in FAD ReN-mAP cells (Extended Data Fig. 5h). These data strongly suggest that p-tau accumulations in FAD ReN cells are induced by amyloid- β accumulation.

To enhance amyloid- β and p-tau pathologies, ReN-mAP cells with the top 1–2% mCherry signals were enriched by FACS (Extended Data Fig. 6a, b). Inhibitors of β - or γ -secretase decreased A β_{40} and A β_{42} levels in these cells while the γ -secretase modulator SGSM41 preferentially decreased toxic A β_{42} levels and increased A β_{38} levels (Extended Data Fig. 6c)¹⁹. PHF1 p-tau levels were also dramatically elevated in the enriched ReN-mAP cells (Fig. 4a, b). PHF1/MAP2 (anti-p-tau/anti-microtubule associated protein 2) and AT8/Tuj1 (anti-p-tau/anti- β -tubulin type III) co-immunofluorescence analysis showed that p-tau levels were largely increased in neurites and neuronal cell bodies (Fig. 4a and Extended Data Fig. 7a). IHC analysis confirmed the dramatic increases in PHF1 p-tau levels in the enriched ReN-mAP cells (Fig. 4b). Notably, treatment with β - or γ -secretase inhibitors reduced the elevated p-tau levels, especially in neurite-like structures (Fig. 4b). Western blot analysis also showed that p-tau levels were robustly increased

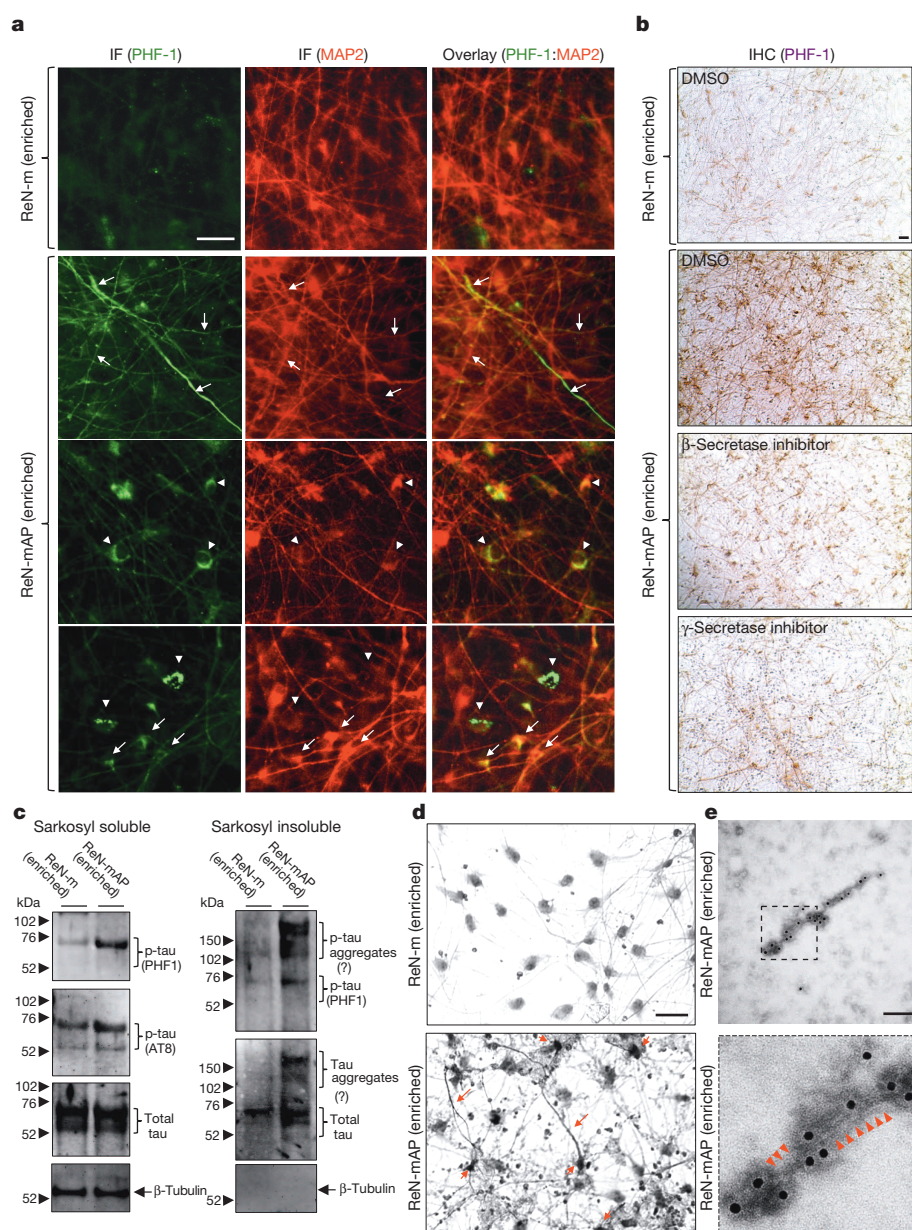


Figure 4 | Detection of aggregated p-tau in the enriched ReN-mAP cells. **a**, Immunofluorescence of p-tau and MAP2 in the enriched ReN-mAP and ReN-m cells after 3D differentiation. Green, p-tau (PHF1); red (pseudo-coloured), MAP2; scale bar, 25 μ m; arrows, p-tau-positive neurites; arrowheads, p-tau-positive cell bodies. **b**, IHC of p-tau in the enriched ReN-mAP and ReN-m cells after 10-week 3D differentiation. The cells were treated with 1 μ M β -secretase inhibitor IV, 3.7 nM compound E or DMSO during the final 2 weeks of the 3D differentiation. PHF1 antibody detected the elevated levels of p-tau in soma and neurites. Brown, p-tau; scale bar, 25 μ m. **c**, Western blot of total and p-tau levels in 1% sarkosyl-soluble and -insoluble fractions. **d**, The modified Gallyas silver staining showed robust increases of strong silver deposits in cell bodies and neurite-like structures in the enriched ReN-mAP cells (lower panel) but not in the enriched ReN-m cells (upper panel). Scale bar, 25 μ m; arrows indicate strong silver deposits in soma and neurites. **e**, Tau filaments were detected in sarkosyl-insoluble fractions from the enriched ReN-mAP cells by transmission electron microscopy after 7-week 3D differentiation. Black dots, anti-tau (τ_{46}) antibodies labelled with immunogold anti-mouse antibodies; scale bar, 100 nm. Lower panel shows a digitally enlarged image (~ 3.5 fold) (arrowheads indicate filamentous structures).

in the enriched ReN-mAP cells while total tau levels were unchanged (Extended Data Fig. 7b). A β -secretase inhibitor again decreased p-tau levels without affecting total tau levels (Extended Data Fig. 7b).

We further explored whether elevated p-tau proteins are aggregated in a manner similar to those observed in the degenerating neurons of Alzheimer's disease. Levels of p-tau were markedly increased in sarkosyl-insoluble fractions of the enriched ReN-mAP cells, but not in control ReN-m cells (Fig. 4c)²¹. The high-molecular-mass p-tau bands were detected only in sarkosyl-insoluble fractions of the enriched ReN-mAP cells, suggesting the presence of sarkosyl/SDS-resistant tau aggregates (Fig. 4c). Next, we performed modified Gallyas silver staining²² on cells that had undergone 10 weeks of differentiation. Notably, strong silver depositions in cell bodies and neurite-like structures were detected in the enriched ReN-mAP, but not in the enriched ReN-m control cells (Fig. 4d, arrows). Finally, immunoelectron microscopy was employed to search for filamentous assemblies of tau protein in sarkosyl-insoluble fractions. Sarkosyl-insoluble fractions from 7-week differentiated ReN-mAP (enriched) and control ReN-G cells were analysed using transmission electron microscopy. Filamentous structures labelled with an anti-tau (τ_{46}) antibody were detected in ReN-mAP cells, not in ReN-G cells (Fig. 4e and Extended Data Fig. 8a, b). The resulting p-tau

aggregates shared striking similarity with those observed in the brains of Alzheimer's disease patients²⁰. We also found that the GSK3 β inhibitors 1-azakenpaullone (1-AZA (ref. 23); Extended Data Fig. 9a, b) and SB415286 (SB41 (ref. 24), data not shown) dramatically decreased p-tau levels in the enriched ReN-mAP cells without changing total tau levels or significantly affecting amyloid- β levels (Extended Data Fig. 9c). These data indicate that GSK3 β regulates amyloid- β -induced tauopathy downstream of amyloid- β deposition^{25,26}.

While increased p-tau levels have been reported in several Alzheimer's disease mouse models with FAD mutations, no somatic accumulated aggregates, detergent-resistant/silver-positive p-tau aggregates, nor immunogold-labelled detergent-resistant tau fibrils were observed^{4,5}. The discrepancies between mouse FAD models and our 3D culture model might be a result of high levels of amyloid- β toxicity that can only be achieved in *in vitro* 3D culture conditions, or the differential tau gene structures in humans. We have shown that 3D-differentiated ReN cells exhibited a dramatic increase in mature human 4-repeat tau isoforms, which may be important for reconstituting tauopathy (Extended Data Fig. 2d). Indeed, a recent study showed that a rat FAD model, which has six tau isoforms similar to human, displayed some aspects of tauopathy²⁷. Moreover, all aspects of tauopathy in our FAD hNPC models were

dramatically attenuated by β - or γ -secretase inhibitors, most probably through the inhibition of amyloid- β generation. These data support that tauopathy is driven by the excessive accumulation of amyloid- β engendered by FAD mutations in our model.

In summary, we have successfully recapitulated amyloid- β and tau pathologies in a 3D human neural cell culture system, which can be used as a platform for studying the pathogenic mechanisms of Alzheimer's disease and drug screening. Our 3D neural cell culture model also provides a unique platform to explore the molecular mechanisms by which p-tau pathologies are induced by toxic amyloid- β species in the absence of frontotemporal lobar degenerative tau mutations. Most importantly, we provide experimental validation of the amyloid hypothesis of Alzheimer's disease, which proposes that the accumulation of amyloid- β drives tauopathy. Our unique strategy for recapitulating Alzheimer's disease pathology in the 3D human neural cell culture model may also serve to facilitate the development of more precise human cellular models of other neurodegenerative disorders.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 January; accepted 26 August 2014.

Published online 12 October 2014.

1. Tanzi, R. E. & Bertram, L. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell* **120**, 545–555 (2005).
2. Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).
3. Selkoe, D. Alzheimer's disease is a synaptic failure. *Science* **298**, 789–791 (2002).
4. Duff, K. Transgenic mouse models of Alzheimer's disease: phenotype and mechanisms of pathogenesis. *Biochem. Soc. Symp.* **67**, 195–202 (2001).
5. Chin, J. Selecting a mouse model of Alzheimer's disease. *Methods Mol. Biol.* **670**, 169–189 (2011).
6. Choi, S. H. & Tanzi, R. E. iPSCs to the rescue in Alzheimer's research. *Cell Stem Cell* **10**, 235–236 (2012).
7. Yagi, T. *et al.* Modeling familial Alzheimer's disease with induced pluripotent stem cells. *Hum. Mol. Genet.* **20**, 4530–4539 (2011).
8. Israel, M. A. *et al.* Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* **482**, 216–220 (2012).
9. Kondo, T. *et al.* Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular A β and differential drug responsiveness. *Cell Stem Cell* **12**, 487–496 (2013).
10. Muratore, C. R. *et al.* The familial Alzheimer's disease APPV717I mutation alters APP processing and Tau expression in iPSC-derived neurons. *Hum. Mol. Genet.* **23**, 3523–3536 (2014).
11. Sproul, A. A. *et al.* Characterization and molecular profiling of PSEN1 familial Alzheimer's disease iPSC-derived neural progenitors. *PLoS ONE* **9**, e84547 (2014).
12. Donato, R. *et al.* Differential development of neuronal physiological responsiveness in two human neural stem cell lines. *BMC Neurosci.* **8**, 36 (2007).
13. Morgan, P. J. *et al.* Protection of neurons derived from human neural progenitor cells by veratridine. *Neuroreport* **20**, 1225–1229 (2009).
14. Mazemondet, O. *et al.* Quantitative and kinetic profile of Wnt/ β -catenin signaling components during human neural progenitor cell differentiation. *Cell. Mol. Biol. Lett.* **16**, 515–538 (2011).
15. Li, H., Wijekoon, A. & Leipzig, N. D. 3D differentiation of neural stem cells in macroporous photopolymerizable hydrogel scaffolds. *PLoS ONE* **7**, e48824 (2012).
16. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
17. Tang-Schomer, M. D. *et al.* Bioengineered functional brain-like cortical tissue. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1324214111> (August 11 2014).
18. Higuchi, M., Trojanowski, J. Q. & Lee, V. M.-Y. Tau protein and tauopathy. *Neuropsychopharmacology: The Fifth Generation of Progress* (eds Davis, K. L., Charney, D., Coyle, J. T. & Nemeroff, C.) 1339–1354 (Lippincott, Williams & Wilkins, 2002).
19. Wagner, S. L. *et al.* Soluble γ -secretase modulators selectively inhibit the production of the 42-amino acid amyloid β peptide variant and augment the production of multiple carboxy-truncated amyloid β species. *Biochemistry* **53**, 702–713 (2014).
20. Trojanowski, J. Q. & Lee, V. M.-Y. The role of tau in Alzheimer's disease. *Med. Clin. North Am.* **86**, 615–627 (2002).
21. Cras, P. *et al.* Neuronal and microglial involvement in β -amyloid protein deposition in Alzheimer's disease. *Am. J. Pathol.* **137**, 241–246 (1990).
22. Nadler, J. V. & Evenson, D. A. Use of excitatory amino acids to make axon-sparing lesions of hypothalamus. *Methods Enzymol.* **103**, 393–400 (1983).
23. Kunick, C., Lauenroth, K., Leost, M., Meijer, L. & Lemcke, T. 1-Azakenpallone is a selective inhibitor of glycogen synthase kinase-3 β . *Bioorg. Med. Chem. Lett.* **14**, 413–416 (2004).
24. Kirby, L. A. *et al.* Glycogen synthase kinase 3 (GSK3) inhibitor, SB-216763, promotes pluripotency in mouse embryonic stem cells. *PLoS ONE* **7**, e39329 (2012).
25. Phiel, C. J., Wilson, C. A., Lee, V. M.-Y. & Klein, P. S. GSK-3 α regulates production of Alzheimer's disease amyloid- β peptides. *Nature* **423**, 435–439 (2003).
26. Jaworski, T. *et al.* GSK-3 α / β kinases and amyloid production *in vivo*. *Nature* **423**, 435–439 (2003).
27. Cohen, R. M. *et al.* A transgenic Alzheimer rat with plaques, tau pathology, behavioral impairment, oligomeric A β , and frank neuronal loss. *J. Neurosci.* **33**, 6245–6256 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the grants from the Cure Alzheimer's fund (D.Y.K., S.H.C. and R.E.T.) and national Institute of Health grants 5P01AG15379 (R.E.T.) and 5R37MH060009 (R.E.T.). We thank T. L. Spires, M. Polydoro and S. Wegmann for revising the manuscript, and M. L. McKee for the electron microscopy assistance. We also appreciate B. T. Hyman, O. Berezovska, J. Hardy and P. Davies for providing cDNAs and antibodies. We acknowledge Ragon Institute's Imaging Core facility (part of the Harvard CFAR Immunology Core), Massachusetts General Hospital (MGH) Viral Vector Core (supported by NIH/NINDS P30NS04776), MGH Microscopy Core of the Center for Systems Biology for immunoelectron microscopy (partially supported by an IBDG Grant DK43351 and a BADERC Award DK57521), MGH Confocal Microscope Core and MGH Pathology Core for technical and instrument support.

Author Contributions D.Y.K. and R.E.T. were equally responsible for experimental design and data interpretation. S.H.C., Y.H.K. and D.Y.K. mainly contributed to writing and revising the manuscript. D.Y.K., Y.H.K., S.H.C., M.H., S.L., C.D., H.C., C.S., B.H., J.B.K., C.A. and C.Z. conducted the experiments. S.L.W. synthesized SGSM41 and S.L.W. and C.Z. characterized SGSM41. J.M., B.J.W., M.P., C.J.W. and D.M.K. contributed to data interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.Y.K. (dkim@helix.mgh.harvard.edu) or R.E.T. (tanzi@helix.mgh.harvard.edu).

METHODS

Cell lines, media and reagents. ReNcell VM human neural precursor (ReN) cells were purchased from EMD Millipore (Billerica, MA, USA). The cells were plated onto BD Matrigel (BD Biosciences, San Jose, CA, USA)-coated T25 cell culture flasks (BD Biosciences, San Jose, CA, USA) and maintained in DMEM/F12 (Life Technologies, Grand Island, NY, USA) media supplemented with $2 \mu\text{g ml}^{-1}$ heparin (StemCell Technologies, Vancouver, Canada), 2% (v/v) B27 neural supplement (Life Technologies, Grand Island, NY, USA), $20 \mu\text{g ml}^{-1}$ EGF (Sigma-Aldrich, St Louis, MO, USA), $20 \mu\text{g ml}^{-1}$ bFGF (Stemgent, Cambridge, MA, USA) and 1% (v/v) penicillin/streptomycin/amphotericin-b solution (Lonza, Hopkinton, MA, USA) in a CO₂ cell culture incubator. The cell culture media were changed every 3 days until the cells were confluent. For 2D neuronal/glia differentiation, the cells were plated onto either Matrigel-coated 24-well or 6-well plates with DMEM/F12 differentiation media supplemented with $2 \mu\text{g ml}^{-1}$ heparin, 2% (v/v) B27 neural supplement, and 1% (v/v) penicillin/streptomycin/amphotericin-b solution without growth factors. One half volume of the differentiation media was changed every 3 days for 3–7 weeks. DAPT, compound E and β -secretase inhibitor IV were purchased from EMD Millipore; *N*-lauroylsarcosine (Sarkosyl) from Sigma-Aldrich; haematoxylin QS from Vector Laboratories (Burlingame, CA, USA); and Amylo-Glo from Biosensis (Thebarton, Australia). SGSM41 is an aminothiazole-bridged heterocycle-containing soluble γ -secretase modulator (SGSM) similar in structure to those published recently¹⁹. SGSM41 has the typical characteristics of this series of SGSM molecules that potentially inhibit the production of A β_{42} and to a lesser degree A β_{40} , while concomitantly increasing the generation of shorter amyloid- β peptide species such as A β_{38} and A β_{37} . The structure and the detailed properties are included in Extended Data Table 1.

Complementary DNA constructs and viral packaging. The construct encoding full-length human β -amyloid precursor protein (APP₆₉₅) with the V717I (London) mutation was obtained from O. Berezovska (Massachusetts General Hospital, Boston, MA, USA; GenBank accession no. NM_201414). The human presenilin 1 (PSEN1) construct with Δ E9 mutation was a gift from J. Hardy (NIH, Bethesda, MD, USA; GenBank accession no. NM_000021). To introduce K670N/M671L (Swedish) mutations into the APP₆₉₅ (London) gene, we performed a site-directed mutagenesis using a mutagenic primer set, 5'-CGGAGGAGATCTCTGAAGTGAATTTGGA TGCAGAATTCGA-3' and 5'-TCGGAATTCTGCATCCAAATTCACCTTCAG AGATCTCTCC G-3' by using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). APP (Swedish/London) and/or PSEN1(Δ E9) cDNAs were then PCR-amplified with Pfu (New England Biolabs, Ipswich, MA, USA) and cloned into lentiviral polycistronic CSCW-IRES-GFP or CSCW-IRES-mCherry vectors to generate CSCW-APP-GFP, CSCW-PSEN1(Δ E9)-mCherry and CSCW-APP-IRES-PSEN1(Δ E9)-IRES-mCherry. The parental CSCW-IRES vector was provided by Massachusetts General Hospital (MGH) viral core. The primers used for the cloning were: APP, 5'-CACCGCTAGCCAGGGTCGCGA ATGCTGC-3' and 5'-GGCGTCGACCTAGTTCTGCATCTGCTC-3'; PSEN1, 5'-CACCGCTAGCAGTTGCTCCAATGACAGAGTTAC-3' and 5'-GACCTCGAG CTAGATATAAAATTGATGAATGC-3'. The amplified APP or PSEN1 genes were double-digested with NheI/SaI or NheI/XhoI and ligated to either CSCW-IRES-GFP or CSCW-IRES-mCherry vectors. To make CSCW-APP-IRES-PSEN1(Δ E9)-IRES-mCherry vector, the APP-IRES segment of CSCW-APP-IRES-GFP vector was PCR-amplified with 5'-CACCGCTAGCCAGGGTCGCGAATGCTGC-3' and 5'-GGCGCTAGCGGTTGTGCCATATTATCATC-3' primers, digested with NheI and cloned into the NheI site of CSCW-PSEN1(Δ E9)-mCherry vector. All the newly constructed vectors were confirmed by sequencing (MGH sequence core, Charlestown, MA, USA). Viral packaging and titre determination were performed by MGH viral core (Charlestown, MA, USA).

Viral infection of ReN cell VM. To transfect the ReN cells with the lentiviral constructs, 50–100 μl viral solution (1×10^6 TU ml⁻¹) were added to 85% confluent proliferating ReN cells in 6-well dishes, incubated for 24 h, and washed three times to stop the infection. The expression of the infected genes was confirmed by mCherry or GFP expression by fluorescence microscopy and western blot analysis.

FACS enrichment of the transfected ReN cells. The infected ReN cells were washed with PBS and then incubated with Accutase (Millipore) for 5 min. The cell pellets were resuspended in PBS supplemented with 2% serum replacement solution (Life Technologies) and 2% B27, and then passed through a cell strainer filter (70 μm Nylon, BD Biosciences). The cell concentrations were adjusted to 2×10^6 cells per ml and then enriched by using FACS Aria cell sorter (Broad Institute of MIT and Harvard, Cambridge, MA, USA). GFP and/or mCherry channels were used to detect the expression of the transfected genes in the individual cells. The sorted/enriched cells were maintained in normal proliferation media.

Amyloid- β ELISA. A β_{40} and A β_{42} levels were mainly measured by Human/Rat amyloid- β ELISA Kit from Wako (Osaka, Japan). The conditioned media from undifferentiated or differentiated ReN cells were collected and diluted by 1:10 or 1:100 with a dilution buffer provided by the manufacturer. A Synergy 2 ELISA plate reader

(BioTek, Winooski, VT, USA) was used to quantify A β_{40} and A β_{42} ELISA signals. To simultaneously measure A β_{38} , A β_{40} and A β_{42} levels, a multi-array electrochemiluminescence ELISA kit was used (Meso Scale Discovery, Rockville, MD, USA). **APOE genotyping.** Genomic DNA (5 ng) isolated from undifferentiated ReN-m and ReN-mAP cells was PCR amplified using TaqMan probes for the two APOE SNP markers (rs429358, catalogue no. C_3084793_20 and rs7412, catalogue no. C_904973_10) using TaqMan Universal Master Mix II (Life Technologies). DNA samples of known APOE ϵ 4 genotypes (in duplicate) were used as controls to obtain the genotype clusters, and all the samples were PCR amplified on a CFX384 thermal cycler (Bio-Rad, Hercules CA, USA).

3D cell cultures and differentiation. For thin-layer 3D cultures, BD Matrigel stock solution (BD Biosciences) was diluted with ice-cold ReN cell differentiation medium (1:15 dilution ratio) and then vortexed with the cell pellets for 20 s. The final cell concentration for the mixture was approximately 2×10^6 cells per ml. The cell/Matrigel mixtures were immediately transferred into either Optilux Black/Clear bottom 96-well plates (100 μl in each well, BD Biosciences) or 8-chamber well Lab-Tek II coverglass plates (200 μl in each well, Thermo Scientific, Rockford, IL, USA) using pre-chilled pipettes. The plates were incubated for 1 h at 37 °C to form thin-layer (100–300 μm) 3D gels at the bottom of the plates and the media were changed. The 3D-plated cells were differentiated for 4–12 weeks depending on the experiments; media was changed every 3–4 days. For thick-layer 3D cultures, BD Matrigel solution was diluted with the same volume of the ice-cold ReN cell differentiation medium (1:2 dilution ratio) and vortexed with ReN cell pellets for 20 s. The final cell concentration for the mixture was approximately 1×10^7 cells per ml. Four-hundred microlitres of the cell/Matrigel mixtures were immediately transferred into tissue culture inserts (ThinCerts, 0.4 μm pore size, Greiner Bio-One, Monroe, NC, USA) and then placed in 24-well plates (BD Biosciences). After 1 h incubation at 37 °C, 1 ml of the pre-warmed differentiation media was added and the cultures were maintained for 4–12 weeks; media was changed every 3–4 days. For drug treatments, differentiation media containing either 1 μM β -secretase inhibitor IV, 1 μM DAPT, 3.7 nM compound E or the same volumes of DMSO were added to 4–6-week differentiated 3D-cultured ReN cells and then maintained for an additional 2–3 weeks. The cells were either fixed with 4% paraformaldehyde (PFA) or harvested for extraction and western blot analysis.

Paraffin embedding and sectioning of thick-layer 3D cultures. For paraffin embedding, 3D thick layer cultures were fixed with 4% PFA at room temperature overnight. The PFA-fixed Matrigel was then transferred to a plastic Tissue-Tek Cryomold (Sakura Finetek, Torrance, CA, USA), preloaded with 60 °C liquefied HistoGel (Thermo Scientific). After positioning the fixed 3D Matrigel at the centre, the whole Cryomold with HistoGel/Matrigel was transferred on ice and then incubated for 15 min until the HistoGel was solidified. The HistoGel/Matrigel complex was then further fixed with 4% PFA at 4 °C overnight, washed five times with PBS, and sent for paraffin embedding (MGH pathology core, Charlestown, MA, USA). The paraffin blocks were then cut into 10 μm sections (Leica SM210R sliding microtome, Leica Microsystems Inc., Buffalo Grove, IL, USA), mounted on polylysine-coated glass slides (Thermo Scientific), and incubated at 45 °C overnight. The sections were deparaffinized by two changes of xylene for 5 min each and then serially transferred to 100%, 90%, 70% ethanol solution for 1 min each. The sections were then rinsed with distilled water for 5 min. For immunofluorescence and IHC, the antigen retrieval was performed by heating the slides for 30 min in Citrate-EDTA Buffer containing 10 mM citric acid (pH 6.2), 2 mM EDTA and 0.05% Tween-20.

Immunofluorescence staining. For immunofluorescence of 3D-cultured ReN cells, thin-layer 3D cultures were fixed with 4% PFA at room temperature for 24 h. The fixed cells were then permeabilized and blocked by incubating with a blocking solution containing 50 mM Tris (pH 7.4), 0.1% Tween-20, 4% donkey serum, 1% BSA, 0.1% gelatin and 0.3 M glycine at 4 °C for 12 h. After washing with TBS buffer containing 0.1% (v/v) Tween-20 (TBST), the 3D cultures were incubated with primary antibodies in the blocking solution at 4 °C for 24 h. After washing three times with TBST, the cells were then incubated with TBST overnight with gentle rocking at 4 °C and then further incubated with AlexaFluor secondary antibodies (Life Technologies) overnight at 4 °C. To avoid fluorescence quenching, a drop of anti-fade gold (Life Technologies) was added on top of the fixed/stained thin-layer 3D cultures before imaging. The fluorescence images were captured by an Olympus DSU confocal microscope (Olympus USA, Center Valley, PA, USA) and the image analysis and 3D reconstruction were performed with ImageJ (a public domain image analysis software), IPLabs (BioVision Technologies, Exton, PA, USA) and MetaMorph (Molecular Devices, Sunnyvale, CA, USA) software. The following antibody dilution rates were used in this study: 3D6 anti-amyloid- β antibody (1:500, a gift from Lilly); anti- β -tubulin type III (Tuj1, 1:200, Abcam, Cambridge, MA, USA); anti-GFAP antibody (1:2,000, DAKO, Carpinteria, CA, USA); 1:200, Antibodies Incorporated, Davis, CA, USA); AT8 anti-p-tau antibody (1:40, Thermo Scientific); PHF1 anti-phospho tau antibody (1:1,000, courtesy of P. Davies); anti-tau₄₆ antibody (1:200, Cell Signaling Technology); anti-GluR2 (1:100, Antibodies Incorporated); anti-MAP2

antibodies (1:400, Millipore; 1:200, Cell Signaling Technology); anti-tyrosine hydroxylase (1:100, Cell Signaling Technology); anti-NR2B (1:100, Antibodies Incorporated); anti-GABA(B)R2 (1:100, Cell Signaling Technology); AlexaFluor 350/488/568 anti-mouse, -rabbit and -chicken secondary antibodies (1:200, Life Technologies).

Immunohistochemical staining. For IHC, thin-layer 3D cultures were permeabilized and blocked by incubating with the blocking solution at 4 °C for 12 h. To block endogenous peroxidase activities, the cultures were incubated with 0.05% (v/v) H₂O₂ solution in TBS for 5 min at room temperature, washed with TBST three times and incubated with the blocking solution for 2 h at room temperature. After incubating with the primary antibody solutions for 24 h at 4 °C, the cultures were washed five times with TBST and then incubated with ImmPRESS anti-mouse or -rabbit Ig (ImmPRESS Peroxidase Polymer Detection Kit, Vector Laboratories, Burlingame, CA, USA) for 30–60 min. The cultures were washed five times for 10 min each with TBST and developed by using ImmPACT DAB Peroxidase Substrate kit (Vector Laboratories). The following antibodies and dilution rates have been used in this study: BA27 anti-A β ₄₀ antibody horseradish peroxidase (HRP)-conjugate (1:2, Wako Chemicals USA, Richmond, VA, USA); BC05 anti-A β ₄₂ antibody HRP-conjugate (1:2, Wako Chemicals USA); AT8 anti-p-tau antibody (1:40, Thermo Scientific); PHF1 anti-p-tau antibody (1:1,000), anti-MAP2 antibody (1:200, Cell Signaling Technology); ImmPRESS anti-mouse and -rabbit Ig HRP polymer conjugates (1:2, Vector Laboratories).

Differential detergent extraction for western blot. The thick-layer 3D cultures of ReN cells were homogenized with TBS extraction buffer containing 50 mM Tris (pH 7.4), 150 mM NaCl, 1 mM NaVO₃, 1 mM NaF, a protease inhibitor mixture (Roche Molecular Biochemicals), a phosphatase inhibitor cocktail (Thermo Scientific), 2 mM PNT (EMD Millipore) and 1 mM phenylmethylsulphonyl fluoride (PMSF, Sigma-Aldrich) by using a battery-operated spinning homogenizer (MIDSCI, St Louis, MO, USA). After incubation on ice for 10 min, the samples were centrifuged for 1 h at 100,000g to get TBS-soluble fractions. The TBS-insoluble pellets were then resuspended in 2% SDS extraction buffer containing 50 mM Tris (pH 7.4), 150 mM NaCl, 2% SDS, 1% Triton X-100, 1 mM NaVO₃, 1 mM NaF, a protease inhibitor mixture (Roche Molecular Biochemicals), a phosphatase inhibitor cocktail (Thermo Scientific), 2 mM PNT (EMD Millipore) and 1 mM PMSF (Sigma-Aldrich) and incubated on ice for an additional 30 min. The samples were centrifuged for 1 h at 100,000g and the supernatant fractions were collected as TBS-insoluble/2% SDS-soluble fractions. The 2% SDS-insoluble pellets were briefly washed with SDS extraction buffer and then further extracted with 90% formic acid (Sigma-Aldrich) on ice and centrifuged for 1 h at 100,000g to produce TBS-insoluble/2% SDS-insoluble/formic-acid-soluble fractions. The formic acid fractions were enriched by using SpeedVac and neutralized by 2 M Tris-Cl buffer (pH 8.3). Protein levels of SDS-soluble fractions were used to normalize the total protein levels in TBS and formic acid fractions. Purification of sarkosyl-insoluble tau was performed as previously described^{28,29} with modifications. TBS-insoluble pellets were resuspended in 1% sarkosyl/RIPA buffer containing 50 mM Tris (pH 7.4), 150 mM NaCl, 0.5% w/v sodium deoxycholate, 2% v/v NP-40, 1% w/v N-lauroylsarcosine, a protease inhibitor mixture (Roche Molecular Biochemicals), a phosphatase inhibitor cocktail (Thermo Scientific), 1 mM NaF, 1 mM NaVO₃, 2 mM PNT (EMD Millipore) and 1 mM PMSF (Sigma-Aldrich) and incubated on ice for 1 h. The samples were then centrifuged for 1 h at 100,000g and the supernatant fractions were collected as 1% sarkosyl/RIPA-soluble fractions. The insoluble pellets were briefly washed two times with 1% sarkosyl/RIPA buffer, further extracted with 90% formic acid (Sigma-Aldrich) on ice and centrifuged for 1 h at 100,000g to produce 1% sarkosyl/RIPA-insoluble /formic-acid-soluble fractions.

Western blot analysis. 15–75 µg of protein were resolved on 12% Bis-Tris or 4–12% gradient Bis/Tris gels (Life Technologies) and the proteins were transferred to nylon membranes (Bio-Rad). For amyloid- β western blot analysis, the membranes were crosslinked with 0.5% glutaraldehyde solution before blocking. Western blot images were visualized by enhanced chemiluminescence (ECL). The images were captured by using BioMax film (Kodak, Rochester, NY, USA) or a VersaDoc imaging system (Bio-Rad) and quantitated by using QuantityOne software (Bio-Rad). Primary antibodies were used at the following dilutions: 6E10 anti-amyloid- β (1:300, Convince); anti-PSEN1 (1:1,000, Cell Signaling Technology); anti- α -tubulin (1:1,000, Cell Signaling Technology); anti-CNPase (1:1,000, Cell Signaling Technology); anti- β -secretase 1 (1:1,000, Cell Signaling Technology); C66 APP C-terminal antibody (1:2,000); AT8 anti-p-tau (1:100, Millipore); PHF1 anti-p-tau (1:500); anti-total tau (1:2,000, DAKO); anti-MAP2 (1:500, Millipore; 1:200, Cell Signaling); anti-NCAM (1:1,000, Cell Signaling); anti-synapsin I (1:500, Cell Signaling); anti-HSP70 (1:1,000, Enzo Life Sciences, Farmingdale, NY, USA) and anti-human mitochondrial antigen (1:500, Millipore).

Modified Gallyas silver staining. Gallyas silver staining was performed with a modified protocol described by Nadler *et al.*²². The PFA-fixed thin-layer 3D cultures were washed three times with deionized distilled water (DDW) for 5 min and

then incubated two times with a pretreating solution containing 4.5% (w/v) NaOH and 0.6% (w/v) ammonium nitrate. The cultures were then incubated with an impregnation solution containing 5.4% (w/v) NaOH, 6.4% (w/v) ammonium nitrate and 0.3% (w/v) silver nitrate for 10 min followed by three washes with a washing solution containing 1 ml of 0.0012% (w/v) ammonium nitrate, 0.5% (w/v) sodium carbonate and 28.5% (v/v) ethanol at 5-min intervals. The deposition of silver particles was detected by incubating with a developer solution with 0.012% (w/v) ammonium nitrate, 0.05% (w/v) citric acid, 0.56% (w/v) formalin and 9.5% (v/v) ethanol for 1–5 min. The reaction was stopped by adding 0.5% acetic acid.

Amylo-Glo staining. The fixed 3D thin-layer cultures of ReN cells were washed three times with saline and incubated with 0.001% (v/v) Amylo-Glo solution for 5 min in a dark environment. The cells were then washed with saline followed by three washings with DDW. To avoid fluorescence quenching, a drop of anti-fade gold (Life Technologies) was added on top of the stained cells. The Amylo-Glo fluorescence was measured with an Olympus DSU confocal microscope with Metamorph image analysis software (Olympus).

Congo red staining. Bannhold's Congo red staining protocol was used for staining paraffin sectioned 3D cultures with slight modification. The hydrated paraffin-sectioned 3D cultures were incubated with 1% Congo red solution (Sigma-Aldrich) for 60 min at room temperature. After briefly rinsing with distilled water three times, the sections were dipped several times into an alkaline alcohol solution (30% EtOH, 0.01% (w/v) NaOH) until the background was cleared. The slides were then washed twice with DDW.

RNA extraction, cDNA synthesis and quantitative RT-PCR (qPCR) analysis. Total RNAs were prepared by using RNeasy mini-columns (Qiagen, Valencia, CA, USA) according to the manufacturer's protocols. Complementary DNAs were synthesized by SuperScript III first-strand synthesis kit (Life Technologies). The pre-validated primer sets for neural markers were purchased from Real Time Primers, LLC (Elkins Park, PA, USA). The amplification was done in a final volume of 20 µl under the following conditions: 15 min at 95 °C and then 55 cycles at 95 °C for 10 s, 58 °C for 45 s and 72 °C for 45 s. The sizes of the qPCR products were confirmed by agarose gel electrophoresis. Biorad iCycler was used to determine count values for each sample. Gene expression levels were normalized against β -actin levels in each sample and the fold changes were calculated by setting the expression levels of each gene in undifferentiated control ReN-G cells as 1. The following are the neuronal and glial marker gene names and PCR product sizes: *NCAM1* (neural cell adhesion molecule 1, 174 base pairs (bp)); *SYT5* (synaptotagmin V, 171 bp); *SLC17A7* (also known as *VGLUT1*, solute carrier family 17 (sodium-dependent inorganic phosphate co-transporter) member 7, 162 bp); *GRIN2A* (glutamate receptor, ionotropic, N-methyl D-aspartate 2A, 170 bp); *EAAT3* (solute carrier family 1, member 1 (neuronal glutamate transporter), 165 bp); *ACHE* (acetylcholinesterase, 232 bp); *SLC6A4* (solute carrier family 6 member 4 (neurotransmitter transporter, dopamine), 213 bp); *GABRA1* (γ -aminobutyric acid (GABA) A receptor alpha 1, 165 bp); *MAPT* (microtubule-associated protein tau, 206 bp); *S100 β* (S100 calcium binding protein, 157 bp); *GFAP* (glial fibrillary acidic protein, 183 bp); *EAAT2* (solute carrier family 1, member 2 (EAAT2, glial), 158 bp); *MBP* (myelin basic protein, 183 bp); and *ACTB* (β -actin, 233 bp).

RT-PCR analysis of 3-repeat and 4-repeat tau levels. Relative levels of 3-repeat and 4-repeat tau mRNAs were determined by RT-PCR using the primer sets: forward 5'-AAGTCGCCGTCTCCGCCAAG-3'; reverse 5'-GTCCAGGGACCCAATCTTCCA-3' as previously described³⁰. The PCR amplification was performed in a final volume of 20 µl under the following conditions: 95 °C for 15 min, and then 30 cycles at 94 °C for 30 s, 60 °C for 30 s and 74 °C for 90 s with a final 10 min extension at 74 °C. RT-PCR products were analysed on 2% agarose gel: 4-repeat tau, 381 bp; 3-repeat tau, 288 bp.

Preparation of sarkosyl-insoluble tau fibrils for immunoelectron microscopy. Sarkosyl-insoluble tau fibrils were prepared from FAD-ReN (ReN-mAP (enriched)) cells, which were differentiated in 3D for 7 weeks^{28,29}. 3D-cultured cell pellets were homogenized in one volume of TBS extraction buffer containing 50 mM Tris (pH 7.4), 150 mM NaCl, 1 mM NaVO₃, 1 mM NaF, a protease inhibitor mixture (Roche Molecular Biochemicals), a phosphatase inhibitor cocktail (Thermo Scientific), 2 mM PNT (EMD Millipore) and 1 mM PMSF (Sigma-Aldrich) by using a battery-operated spinning homogenizer (MIDSCI). TBS homogenates were then mixed and homogenized with one volume of 2 \times RIPA buffer containing 50 mM Tris (pH 7.4), 150 mM NaCl, 0.5% w/v sodium deoxycholate, 2% v/v NP-40, a protease inhibitor mixture (Roche Molecular Biochemicals), a phosphatase inhibitor cocktail (Thermo Scientific), 1 mM NaF, 1 mM NaVO₃, 2 mM PNT (EMD Millipore) and 1 mM PMSF (Sigma-Aldrich), and incubated on ice for 15 min. After centrifugation (18,000g for 20 min at 4 °C), 20% w/v N-lauroylsarcosine (sarkosyl) stock solution was added to the RIPA-soluble supernatant fraction to adjust the final sarkosyl concentration to 1%. After incubating at room temperature with gentle rocking for 1 h, RIPA/sarkosyl homogenates were then centrifuged for 1 h at 150,000g. The pellets were then resuspended in 10 µl of PBS (RIPA-soluble/sarkosyl-insoluble fraction). The

RIPA-insoluble pellets were further homogenized in H buffer containing 10 mM Tris (pH 7.4), 1 mM EGTA, 0.8 M NaCl, 10% w/v sucrose, a protease inhibitor mixture (Roche Molecular Biochemicals), a phosphatase inhibitor cocktail (Thermo Scientific), 1 mM NaF, 1 mM NaVO₃, 2 mM PNT (EMD Millipore) and 1 mM PMSF (Sigma-Aldrich) and then centrifuged at 18,000g for 20 min. 20% w/v *N*-lauroylsarcosine (sarkosyl) stock solution was added to adjust the final sarkosyl solution concentration to 1%. After incubating at room temperature with gentle rocking for 1 h, H buffer/sarkosyl homogenates were then centrifuged for 1 h at 150,000g and the pellets were resuspended in 10 μ l of PBS (RIPA-insoluble/sarkosyl-insoluble fraction). Both RIPA-soluble/sarkosyl-insoluble and RIPA-insoluble/sarkosyl-insoluble fractions were used for immunoelectron microscopy. The same protocol was used to enrich sarkosyl-insoluble p-tau aggregates in Alzheimer's disease brain samples.

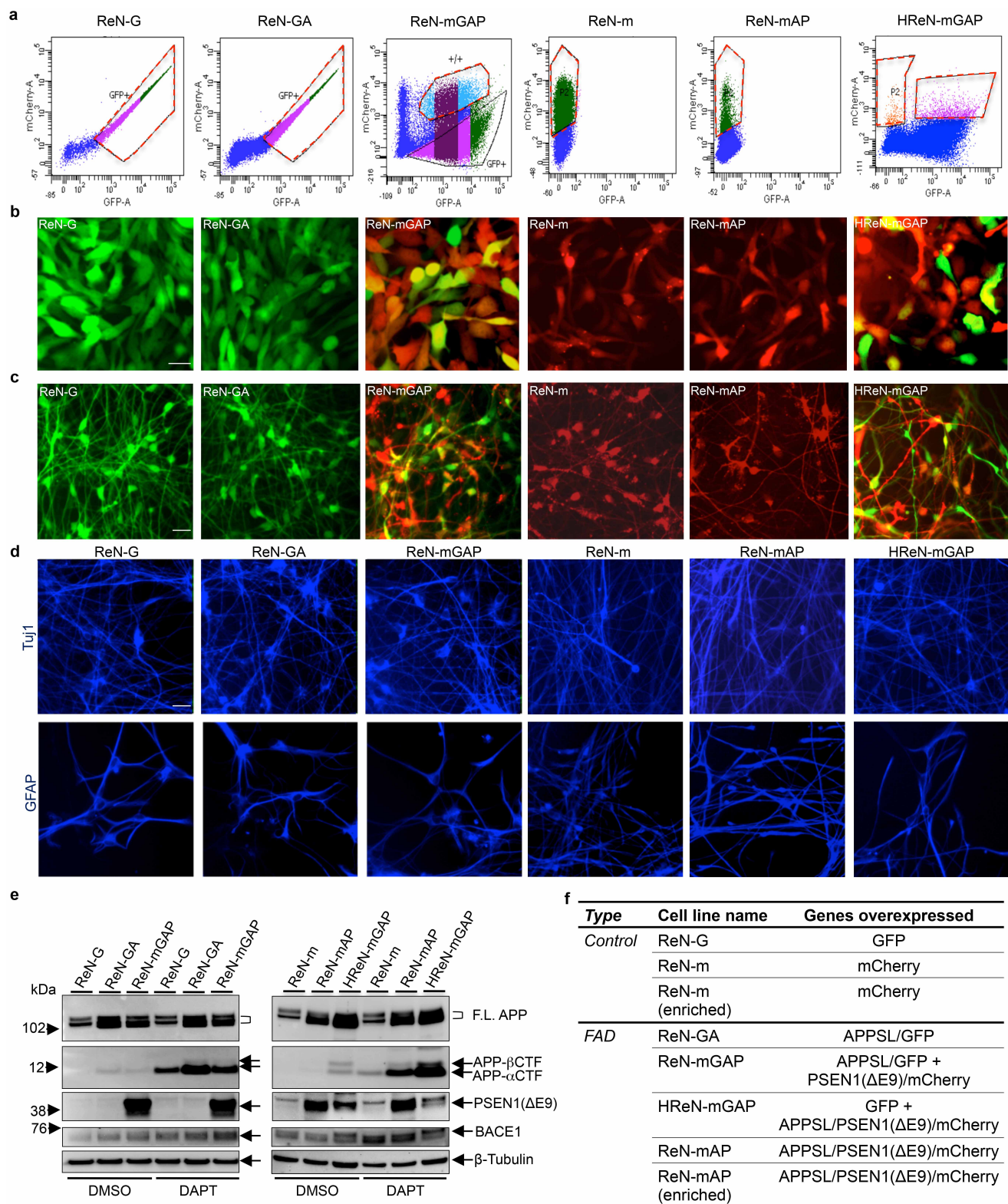
Immunogold staining of sarkosyl-insoluble tau. Immunoelectron microscopy was performed in the Microscopy Core of the Center for Systems Biology/Program in Membrane Biology (MGH, Boston, USA). The sarkosyl-insoluble fractions were resuspended in PBS, placed on formvar-carbon coated Ni grids and allowed to adsorb for 10 min. They were placed on drops of tau₄₆ antibody solution (1:25, Cell Signaling Technology) for 1 h at room temperature, then rinsed on drops of PBS and placed on drops of goat-anti-mouse 10 nm gold (Ted Pella, Redding, CA, USA) for 1 h. They were rinsed on drops of distilled water and stained for 1 min on drops of 2% phosphotungstic acid (Electron Microscopy Sciences, Hatfield, PA, USA). Grids were examined in a JEOL JEM 1011 transmission electron microscope at 80 kV.

Images were collected using an AMT digital imaging system (Advanced Microscopy Techniques, Danvers, MA, USA).

Electrophysiology. Whole-cell recordings were performed on differentiated ReN cells with an Axopatch 200A amplifier (Molecular Devices) and fire polished patch pipettes with resistances of 2–3 M Ω . Pipette solution was 140 mM KCl, 2 mM MgCl₂, 1 mM EGTA, 10 mM HEPES, 4 mM MgATP, 0.3 mM NaGTP, and 0.1 mM Na₂PhosCr, pH 7.2, adjusted with KOH. The external solution was 140 mM NaCl, 5 mM KCl, 2 mM CaCl₂, 2 mM MgCl₂, 10 mM HEPES, and 10 mM D-glucose, pH 7.4, adjusted with NaOH. Command protocols were generated and data were digitized with a Digidata 1440A A/D interface with pCLAMP10 software (Molecular Devices). Voltage errors were minimized by series resistance compensation (80%). During the recording, 500 nM tetrodotoxin was applied to isolate Na⁺ currents by subtracting currents after and before tetrodotoxin application. The data were analysed by Clampfit (Molecular Devices) and Sigmaplot.

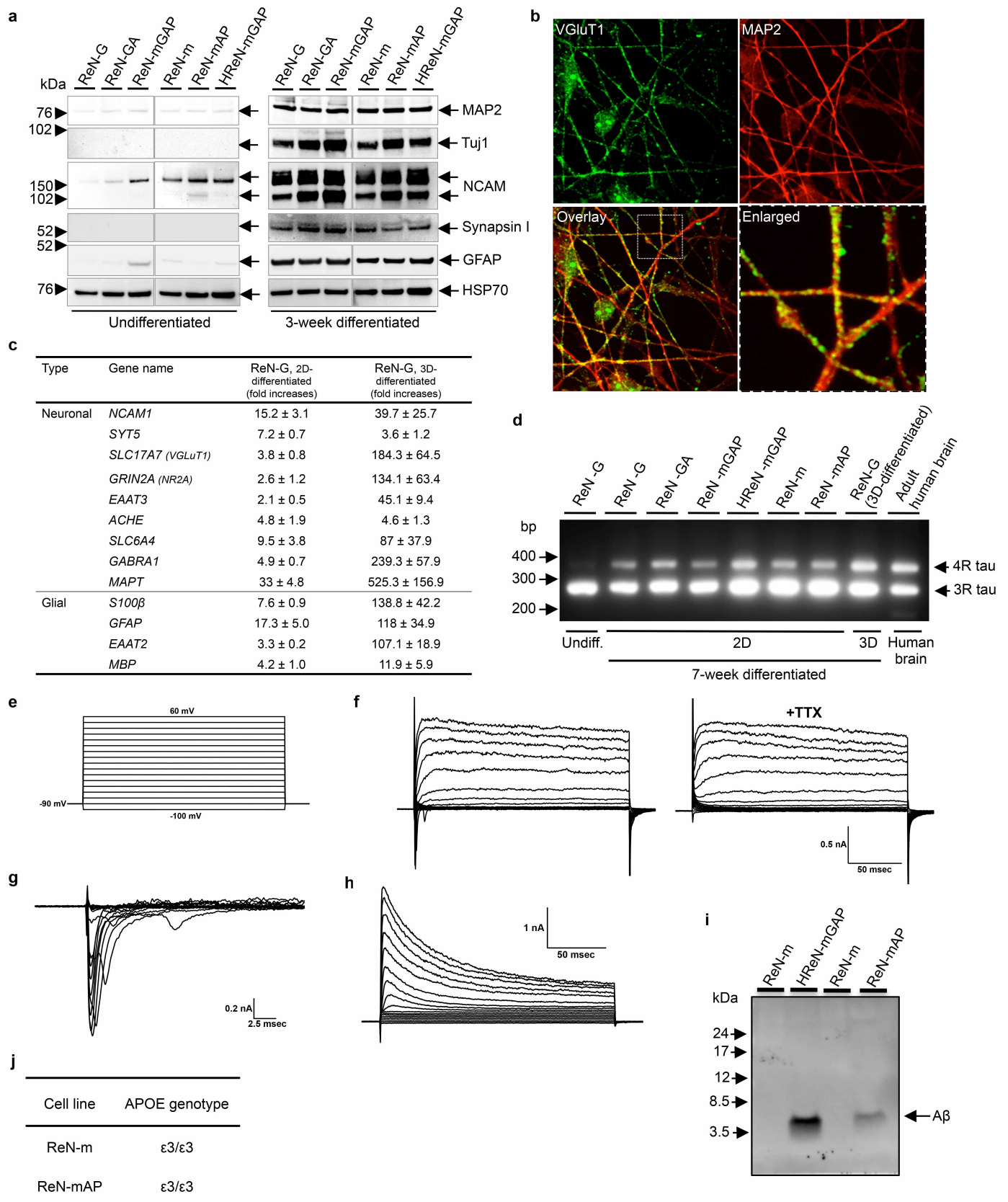
Statistics. All statistical analyses were performed using a two-tailed Student's *t*-test or one-way ANOVA followed by a post hoc Dunnett's test. Data in graphs are expressed as mean values \pm s.e.m. Error bars represent s.e.m.

28. Planel, E. *et al.* Acceleration and persistence of neurofibrillary pathology in a mouse model of tauopathy following anesthesia. *FASEB J.* **23**, 2595–2604 (2009).
29. Julien, C., Bretteville, A. & Planel, E. Biochemical isolation of insoluble tau in transgenic mouse models of tauopathies. *Methods Mol. Biol.* **849**, 473–491 (2012).
30. Iovino, M., Patani, R., Watts, C., Chandran, S. & Spillantini, M. G. Human stem cell-derived neurons: a system to study human tau function and dysfunction. *PLoS ONE* **5**, e13947 (2010).



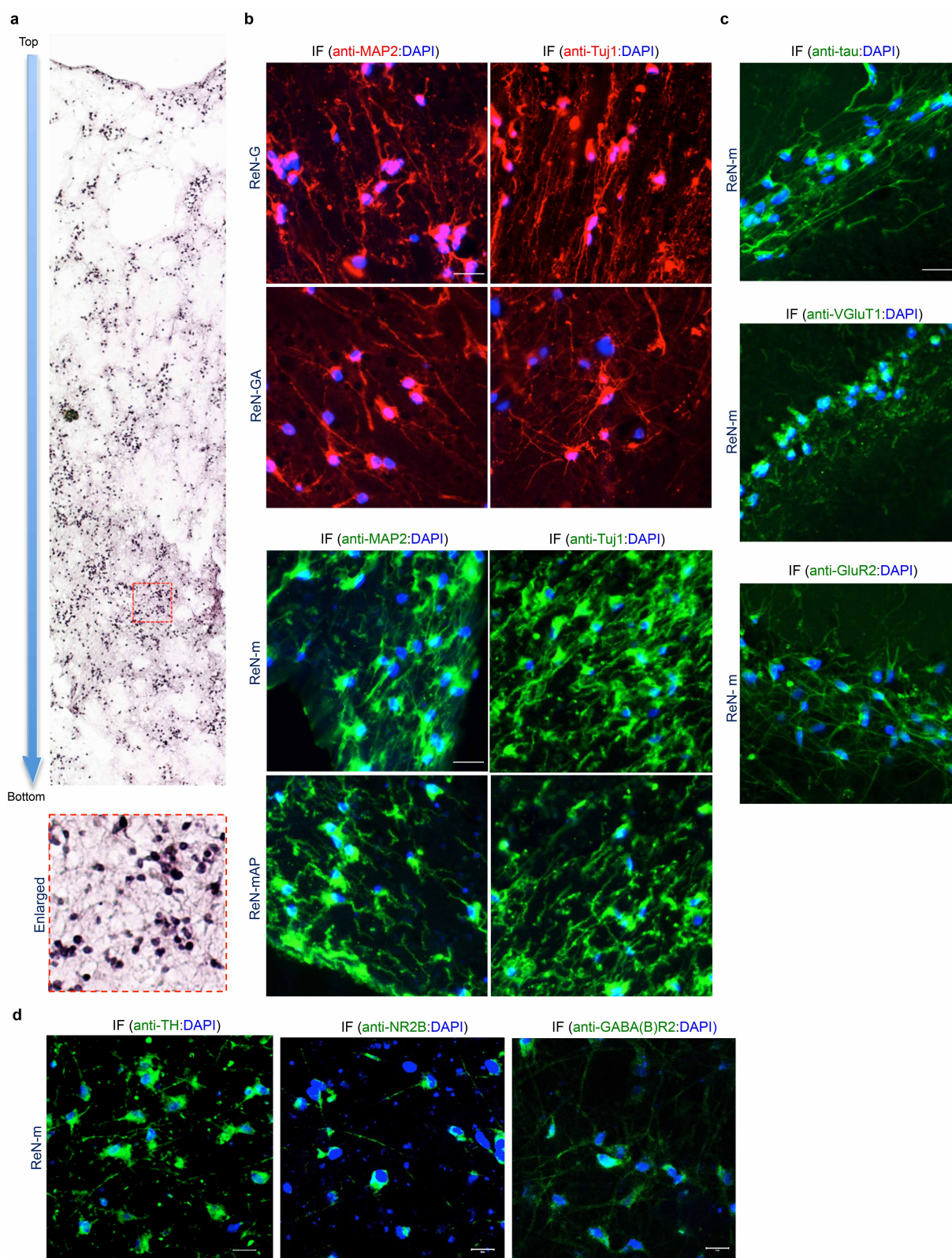
Extended Data Figure 1 | Generation of FACS-sorted ReN cells with FAD mutations. **a**, FACS sorting of ReN cell VM human neural stem (ReN) cells that were stably transfected with polycistronic GFP and/or mCherry lentiviral vector(s). The cells were then enriched based on GFP and/or mCherry signals by FACS (red-dotted boxes, the selected ranges of cells for the experiments; mCherry-A, area intensity of mCherry signal; GFP-A, area intensity of GFP signal). **b**, ReN cells stably expressing GFP alone (ReN-G), APPSL-GFP (ReN-GA), APPSL-GFP-PSEN1(ΔE9)-mCherry (ReN-mGAP), mCherry alone (ReN-m), APPSL-PSEN1(ΔE9)-mCherry (ReN-mAP) or GFP-APPSL-PSEN1(ΔE9)-mCherry (HReN-mGAP). Green, GFP; red, mCherry; scale bar,

25 μm. **c**, The representative fluorescence microscope images of ReN cells that were differentiated by growth-factor deprivation for 3 weeks (green, GFP; red, mCherry; scale bar, 25 μm). **d**, Immunofluorescence of neuronal (Tuj1) and glial markers (GFAP) in 3-week differentiated control and FAD ReN cells. Scale bar, 25 μm. **e**, Western blot of APPSL and PSEN1(ΔE9) expression in control (ReN-G and ReN-m) and FAD ReN (ReN-GA, ReN-mGAP and HReN-mGAP) cells. APP C-terminal fragment (CTF) levels were greatly increased by 500 nM DAPT treatments for 24 h. BACE1, β-secretase 1, F.L. APP, full-length APP. **f**, A table summarizing the control and FAD ReN cells generated for this study.



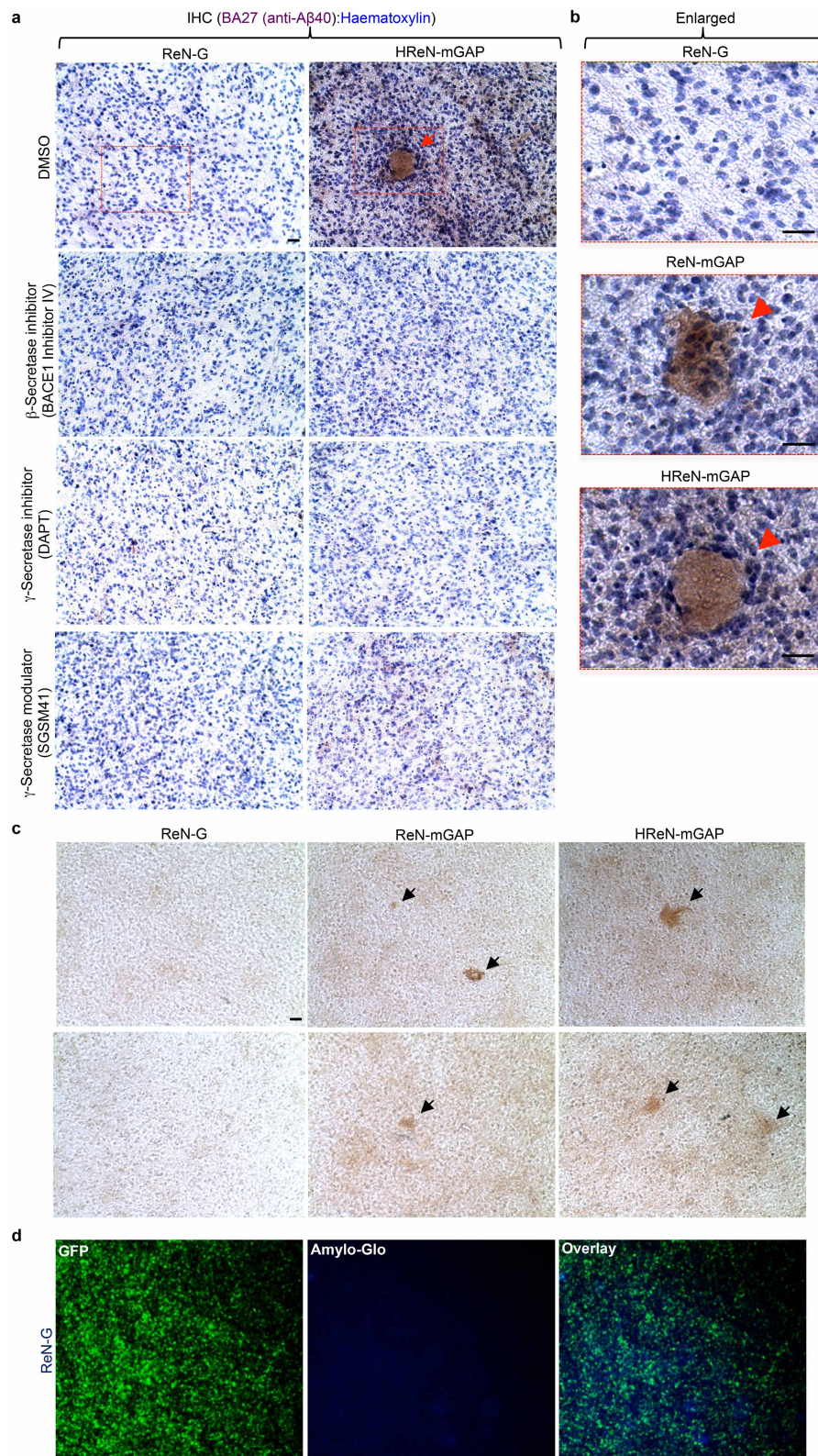
Extended Data Figure 2 | Characterization of the control and FAD ReN cells. **a**, Western blot of neuronal (MAP2, Tuj1, NCAM, synapsin 1) and glial (GFAP) markers in undifferentiated and 3-week differentiated control and FAD ReN cells. **b**, Confocal immunofluorescence of presynaptic (VGluT1, green) and dendritic (MAP2, red (pseudo-coloured)) markers in 6-week differentiated control ReN-m cells. Top-left, top-right and bottom-left panels, $\times 100$ magnification; bottom-right panel represents a digitally magnified image of the respective outlined region for better visualization of punctate structures. **c**, qPCR array analysis of neuronal and glial markers of 7-week differentiated control ReN-G cells. Gene expression levels were normalized against β -actin levels in each sample and the fold changes were calculated by setting the expression levels of each gene in undifferentiated control ReN-G cells as 1 ($n = 3$ for ReN-G whereas $n = 5$ for ReN-G in 3D differentiation). FAD ReN cells (HReN-mGAP and ReN-mAP) showed a similar pattern of increases in neuronal and glial markers (data not shown). **d**, Analysis of 4-repeat (4R) or 3-repeat (3R) tau isoforms in 7-week differentiated control ReN cells (ReN-G and ReN-m) and FAD ReN cells (ReN-GA, ReN-mGAP,

HReN-mGAP and ReN-mAP). Complementary DNA samples prepared from undifferentiated control ReN-G (1st lane) and human adult brains (9th lane) were used as controls. **e**, **f**, Electrophysiological properties of differentiated control ReN-G cells. The currents were elicited by 10 mV voltage steps from -100 to $+60$ mV in external solution (**e**) without (left panel in **f**) or with 500 nM tetrodotoxin (TTX, right panel in **f**). TTX treatment specifically blocked voltage-gated sodium currents. ReN-G cells were differentiated for 29 days by the previously described 'preD' method¹². **g**, Sodium currents are shown as subtracted currents. **h**, Premature ReN-G cells (<16 days) mostly showed voltage-gated potassium currents without TTX-sensitive sodium currents. **i**, Western blot of amyloid- β (A β) levels in the conditioned media collected from 6-week differentiated control (ReN-m) and FAD ReN (ReN-mAP and HReN-mGAP) cells. **j**, A table summarizing APOE genotypes of control (ReN-m) and FAD ReN (ReN-mAP) cells used in this study. Two APOE SNP markers, rs429358 (minor allele = C) and rs7412 (minor allele = T), were used to determine APOE $\epsilon 2/3/4$ genotypes.



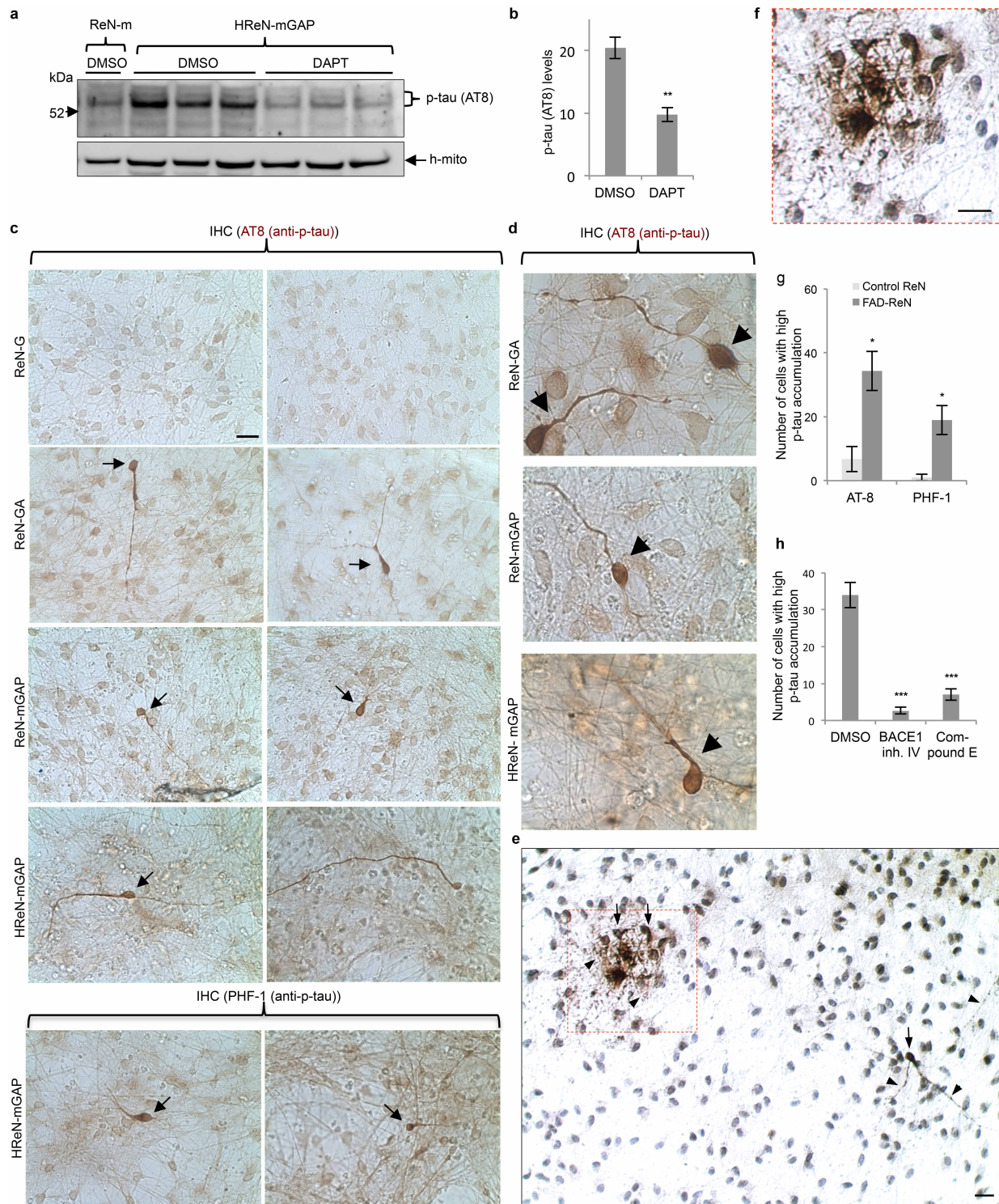
Extended Data Figure 3 | Characterization of differentiated ReN cells in 3D cultures. **a**, Haematoxylin staining of a representative paraffin section (10 μ m) from 9-week differentiated ReN-m cells in thick-layer 3D Matrigel. The sections were vertically cut to show the top and bottom of the 3D cultures. The pictures (10) were serially taken from top to bottom and digitally combined together. Bottom panel represents a digitally magnified image of the respective outlined region for better visualization; scale bar, 50 μ m. **b**, The paraffin sections from control (ReN-G and ReN-m) and FAD ReN (ReN-GA and ReN-mAP (enriched)) were immunofluorescence stained with antibodies

against the neuronal markers Tuj1 and MAP2. Blue, DAPI; scale bar, 25 μ m. **c**, Immunofluorescence of thick-layer 3D-differentiated ReN-m cells stained with antibodies against additional neuronal markers tau, VGlut1 or GluR2 (glutamate receptor, ionotropic, AMPA 2); scale bar, 25 μ m. **d**, Immunofluorescence of 3D-differentiated ReN-m cells (7 weeks) stained with antibodies against mature neuronal markers TH (tyrosine hydroxylase), NR2B (NMDA receptor 2B) or GABA(B)R2 (GABA-B-receptor 2); blue, DAPI; scale bar, 20 μ m.



Extended Data Figure 4 | Reconstitution of amyloid- β aggregates in 3D-cultured FAD ReN cells. **a**, IHC of amyloid- β deposits in control (ReN-G) and FAD ReN (HReN-mGAP) cells differentiated in 3D Matrigel. The control and FAD ReN cells were 3D-differentiated for 6 weeks and then treated with 1 μ M BACE1 inhibitor IV (β -secretase inhibitor), 500 nM DAPT (γ -secretase inhibitor), 500 nM SGSM41 (γ -secretase modulator) or DMSO for an additional 3 weeks. The cultures were then fixed and immunostained with HRP-conjugated BA27 anti-A β_{40} antibodies (brown, DAB (BA27); blue, haematoxylin; scale bars, 25 μ m; arrowhead, large amyloid- β deposits).

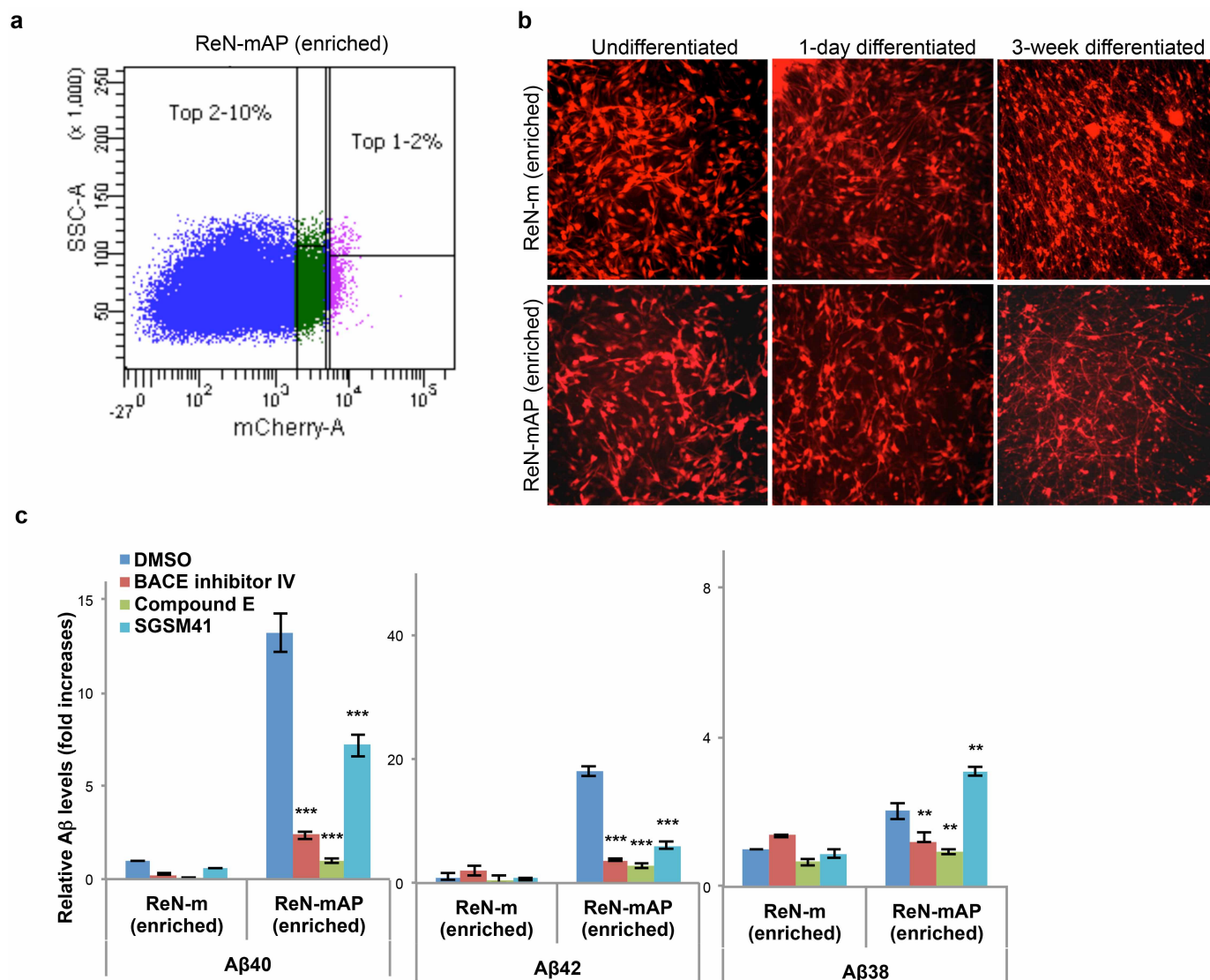
b, Enlarged images of amyloid- β deposits in control (ReN-G) and FAD ReN (ReN-mGAP, HReN-mGAP) cell pictures shown in Fig. 2d and Extended Data Fig. 4a. Scale bar, 25 μ m. **c**, IHC of amyloid- β deposits in control (ReN-G, left panels) and FAD ReN (ReN-mGAP, middle panels; HReN-mGAP, right panels) cells differentiated in 3D thin-layer Matrigels for 9 weeks. The fixed thin-layer 3D cultures were immunostained with HRP-conjugated BA27 anti-A β_{40} antibodies (brown, DAB (BA27); scale bar, 25 μ m; arrows, large amyloid- β deposits). **d**, Amylo-Glo staining of ReN-G 3D cultures. Green, GFP; blue, Amylo-Glo; $\times 10$ magnification.



Extended Data Figure 5 | Accumulation of p-tau in FAD ReN cells.

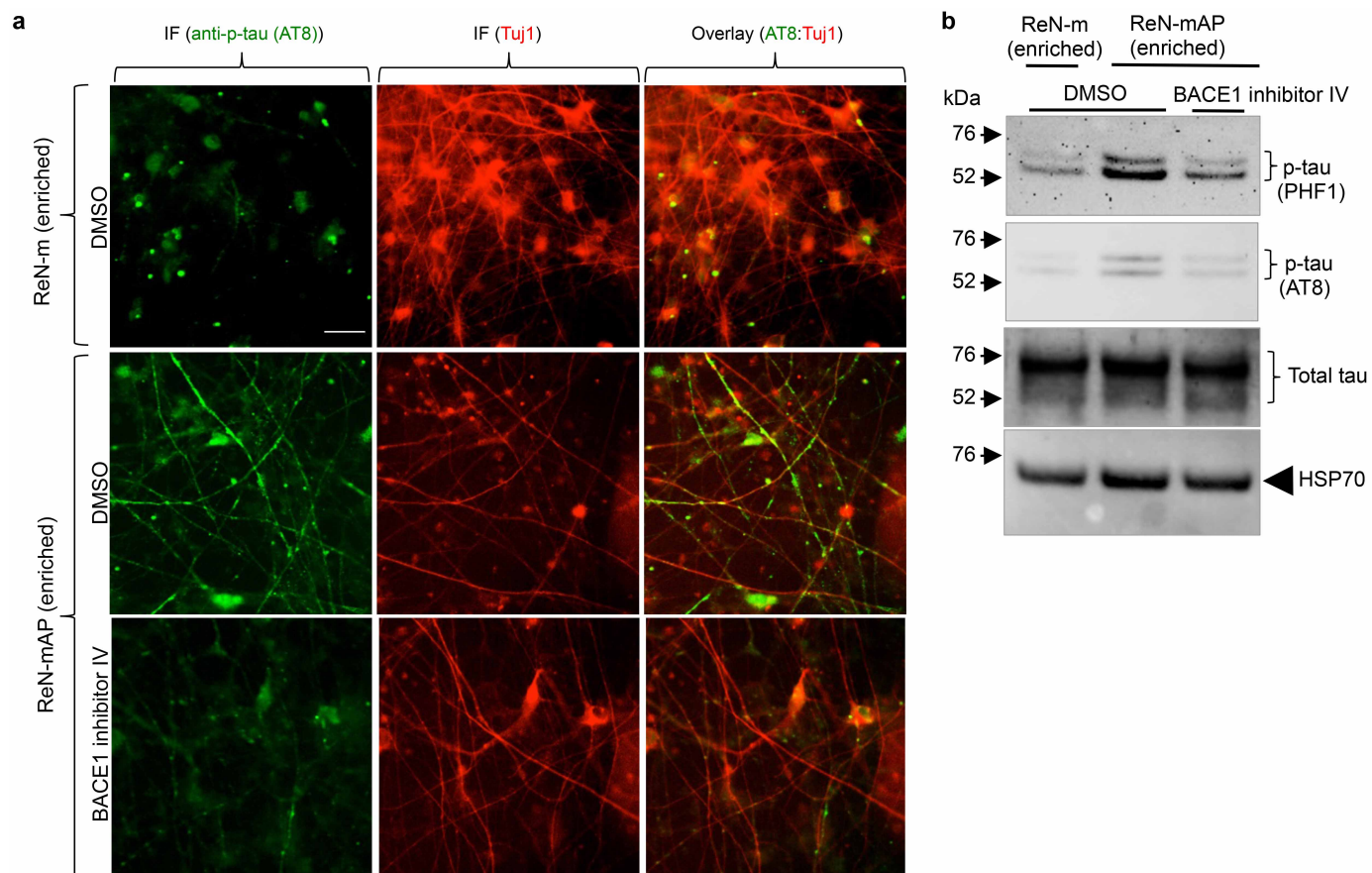
a, Elevated p-tau levels were significantly decreased by 1 μ M DAPT (γ -secretase inhibitor) treatment in 6-week differentiated HReN-mGAP cells. The antibody against human specific mitochondrial marker (h-mito) was used to show an equal loading of the samples. **b**, Quantification of p-tau levels in control (DMSO)- and DAPT-treated HReN-mGAP cells (** $P < 0.01$; t -test; $n = 3$ per each sample). y axis values represent relative signal intensities (% adjusted volumes) of each p-tau band. **c**, p-tau IHC showed p-tau-positive cells in 3D-differentiated FAD ReN cells. Two p-tau antibodies, AT8 (pSer 199/Ser 202/Thr 205) and PHF1 (pSer 396/Ser 404), against different phosphorylation sites were used. Brown, p-tau; scale bar, 25 μ m; arrows indicate cells with high levels of p-tau. **d**, High magnification ($\times 100$) images of p-tau-positive neurons (brown, AT8 p-tau). IHC of AT8 p-tau staining

showed neurons with high levels of p-tau accumulation in soma and neurite-like structures (arrowheads). **e**, IHC of p-tau (AT8) staining showed the cells with p-tau accumulations in soma and neurite-like structures (arrow and arrowhead, respectively); scale bar, 25 μ m. **f**, A digitally enlarged image of a dotted box in e. Brown, p-tau (AT8); blue, haematoxylin; scale bar, 25 μ m. **g**, Total number of cells with high levels of p-tau in a single well of a 96-well plate was counted in control (ReN-G and ReN-m) and FAD ReN (ReN-GA, ReN-mGAP and HReN-mGAP) cells (* $P < 0.05$; t -test; $n = 5$ for control ReN cells and $n = 12$ for FAD ReN cells). **h**, BACE1 inhibitor IV (1 μ M) or compound E (3.7 nM) treatments markedly decreased the number of cells with high p-tau accumulation in ReN-mAP cell populations (*** $P < 0.001$; ANOVA followed by a post hoc Dunnett's test; $n = 3$ for control ReN-m and ReN-mAP cells).



Extended Data Figure 6 | FACS enrichment of ReN-mAP and ReN-m cells for higher expressions of APP and PS1. **a**, FACS sorting of ReN-mAP cells with top 1–2% mCherry signal. mCherry-A, area intensity of mCherry signal. **b**, Representative images of the mCherry-labelled (red), enriched control ReN-m and ReN-mAP cells before or after the differentiation under growth-factor free conditions ($\times 200$ magnification); 1- and 3-week differentiated cells show increased neurite outgrowth. **c**, The enriched ReN-mAP cells secreted high levels of A β ₄₀ and A β ₄₂ after 9-week 3D differentiation.

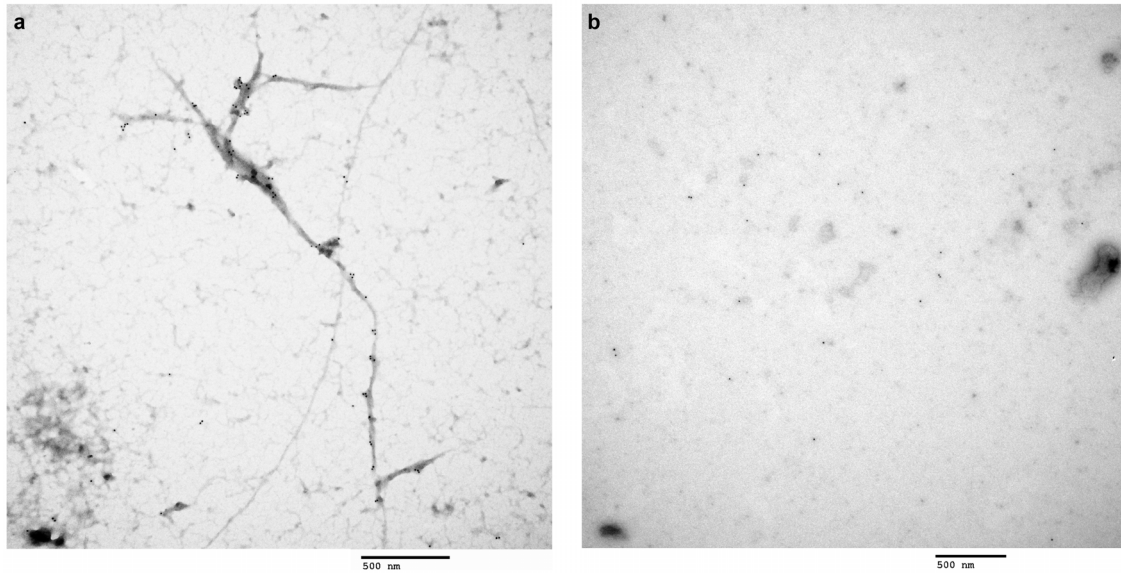
The secreted A β _{38/40/42} levels were measured by a multi-array Meso Scale electrochemiluminescence (Meso Scale SQ 120 system). Relative levels of amyloid- β (fold increases) were calculated by setting amyloid- β levels of the control ReN-m as 1. Amyloid- β levels were changed after treating 1 μ M BACE1 inhibitor IV, 3.7 nM compound E or 500 nM SGSM41 (** P < 0.01; *** P < 0.001; ANOVA followed by a post hoc Dunnett's test; n = 3 for the enriched ReN-m and ReN-mAP cells).



Extended Data Figure 7 | Increased p-tau levels in FAD ReN cells.

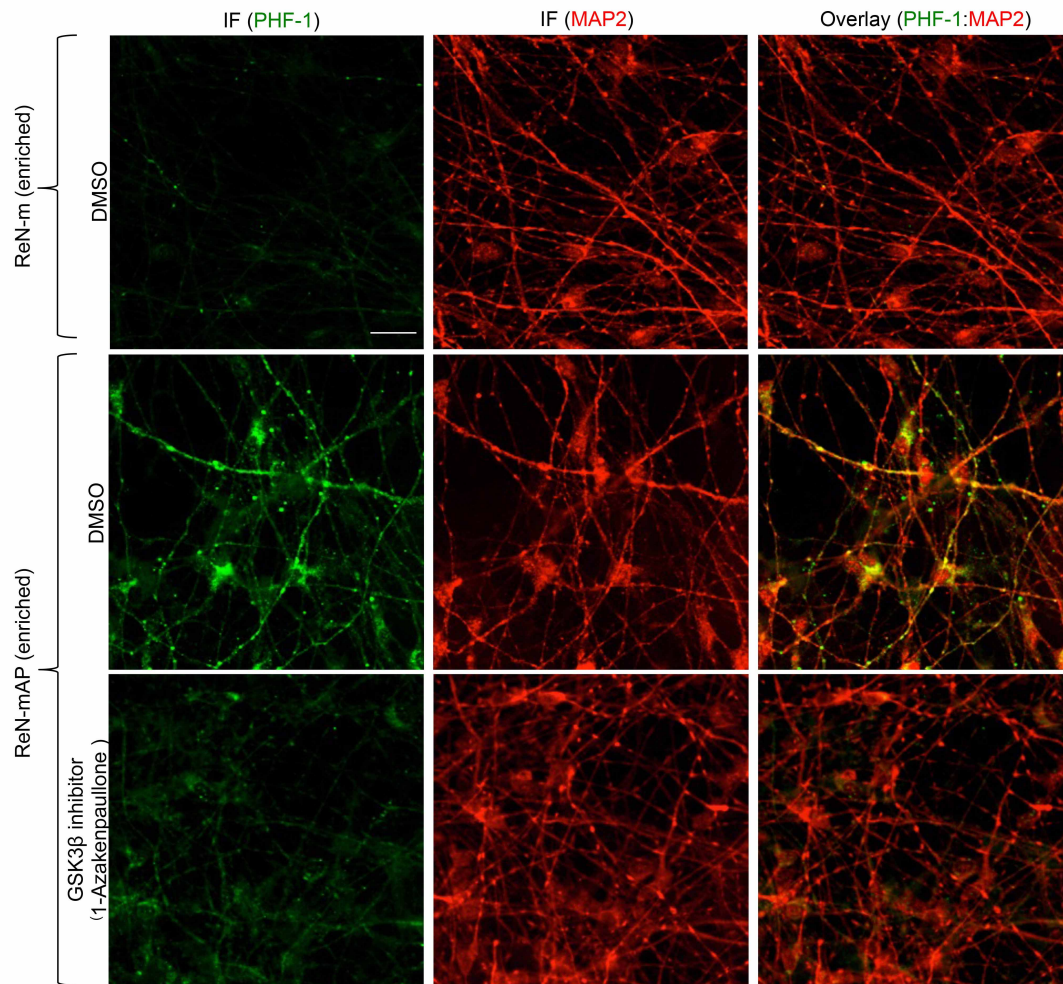
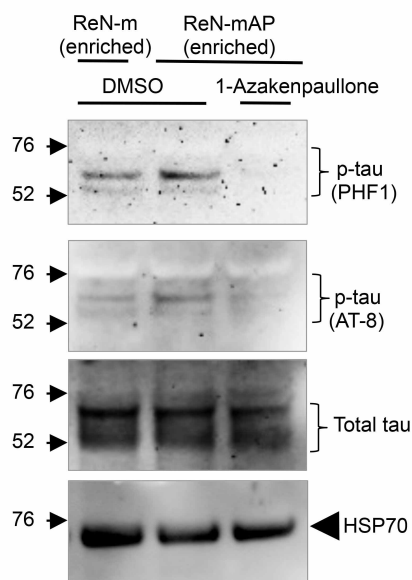
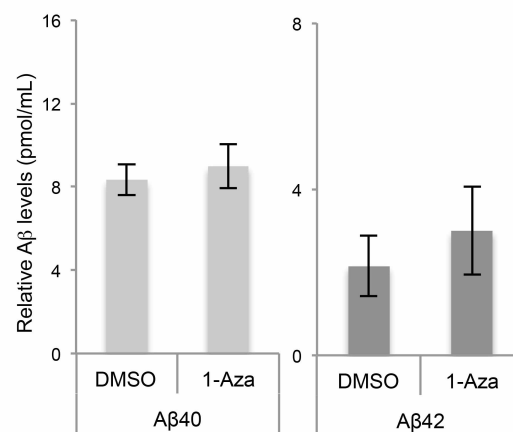
a, Immunofluorescence of AT8 p-tau and Tuj1 in the enriched ReN-mAP and control ReN-m cells after 9 weeks of 3D differentiation. BACE1 inhibitor IV treatment for 3 weeks dramatically reduced AT8 p-tau staining (green, AT8 p-tau; red (pseudo-coloured), Tuj1; scale bar, 25 μ m). **b**, Western blot of total

and p-tau levels in control (enriched ReN-m) and FAD ReN (enriched ReN-mAP) cells. The cells were 3D differentiated for 9 weeks. Three weeks of BACE1 inhibitor treatments significantly decreased p-tau levels without changing total tau levels. HSP70 heat shock protein levels are shown to demonstrate equal loading of each sample.



Extended Data Figure 8 | Immunoelectron microscopy analysis of sarkosyl-insoluble fraction from FAD and control ReN cells. **a**, Sarkosyl-insoluble fractions prepared from 3D-differentiated ReN-mAP (enriched, 7-week differentiated) were placed on carbon grids, labelled with tau₄₆ and anti-mouse 10 nm gold antibodies and imaged using a JEOL JEM 1011 transmission

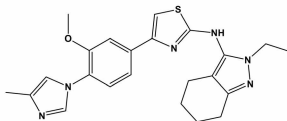
electron microscope (scale bar, 500 nm). **b**, Sarkosyl-insoluble fractions from 3D-differentiated control ReN-G cells (7-week differentiated). No immunogold-labelled filamentous structures were detected in these samples (scale bar, 500 nm).

a**b****c**

Extended Data Figure 9 | Treatment with 1-azakenpaullone, a GSK3 β inhibitor, decreased amyloid- β -induced tau phosphorylation without changing total amyloid- β levels. a, Immunofluorescence of p-tau and MAP2 in the enriched ReN-mAP and control ReN-m cells with or without treatment with 1-azakenpaullone, a GSK3 β inhibitor. The differentiated cells were treated with 2.5 μ M 1-azakenpaullone or DMSO for the last 5 days of the 3D differentiation (green, p-tau (PHF1); red (pseudo-coloured), MAP2;

scale bar, 25 μ m). **b,** Western blot of total and p-tau levels in control (enriched ReN-m) and FAD ReN (enriched ReN-mAP) cells. The cells were 3D differentiated for 4 weeks followed by additional 5-day treatments of DMSO or 2.5 μ M 1-azakenpaullone. **c,** Analysis of A β ₄₀ and A β ₄₂ levels in the enriched ReN-mAP cells treated with either DMSO or 2.5 μ M 1-azakenpaullone (1-Aza) under the same conditions.

Extended Data Table 1 | The structure and the properties of SGSM41, a novel soluble γ -secretase modulator

Compound	Structure	IC ₅₀ for the inhibition of A β ₄₂	IC ₅₀ for the inhibition of A β ₄₀	ClogP	Kinetic solubility (μ M) in PBS pH 7.4	EC ₅₀ for the potentiation of A β ₃₈	Notch	APP	Microsomal Stability % Remaining (human) ; 30 min	Microsomal Stability % Remaining (rat) ; 30 min	Microsomal Stability % Remaining (mouse) ; 30 min
41		115nM	1229nM	4.91	4.6	389nM	NO	NO	28	56	67

SGSM41 has the typical characteristics of this series of SGSM molecules that potently inhibit the production of toxic A β ₄₂, and to a lesser degree A β ₄₀, while concomitantly potentiating the generation of shorter amyloid- β peptide species such as A β ₃₈ and A β ₃₇.

Piezo1 integration of vascular architecture with physiological force

Jing Li^{1*}, Bing Hou^{1*}, Sarka Tumova¹, Katsuhiko Muraki², Alexander Bruns¹, Melanie J. Ludlow¹, Alicia Sedo¹, Adam J. Hyman¹, Lynn McKeown¹, Richard S. Young^{1,3}, Nadira Y. Yuldasheva¹, Yasser Majeed¹, Lesley A. Wilson¹, Baptiste Rode¹, Marc A. Bailey¹, Hyejeong R. Kim⁴, Zhaojun Fu¹, Deborah A. L. Carter¹, Jan Bilton¹, Helen Imrie¹, Paul Ajuh⁵, T. Neil Dear¹, Richard M. Cubbon¹, Mark T. Kearney¹, K. Raj Prasad³, Paul C. Evans⁴, Justin F. X. Ainscough¹ & David J. Beech¹

The mechanisms by which physical forces regulate endothelial cells to determine the complexities of vascular structure and function are enigmatic^{1–5}. Studies of sensory neurons have suggested Piezo proteins as subunits of Ca²⁺-permeable non-selective cationic channels for detection of noxious mechanical impact^{6–8}. Here we show Piezo1 (Fam38a) channels as sensors of frictional force (shear stress) and determinants of vascular structure in both development and adult physiology. Global or endothelial-specific disruption of mouse *Piezo1* profoundly disturbed the developing vasculature and was embryonic lethal within days of the heart beating. Haploinsufficiency was not lethal but endothelial abnormality was detected in mature vessels. The importance of *Piezo1* channels as sensors of blood flow was shown by *Piezo1* dependence of shear-stress-evoked ionic current and calcium influx in endothelial cells and the ability of exogenous *Piezo1* to confer sensitivity to shear stress on otherwise resistant cells. Downstream of this calcium influx there was protease activation and spatial reorganization of endothelial cells to the polarity of the applied force. The data suggest that *Piezo1* channels function as pivotal integrators in vascular biology.

Messenger RNA encoding *Piezo1* was readily detected in mouse aorta and a variety of human endothelial cells (Extended Data Fig. 1). To gain insight into its significance we first investigated cultured human umbilical vein endothelial cells (HUVECs). Depletion of *Piezo1* by either of two short interfering RNAs (siRNAs) strongly suppressed migration of these cells towards vascular endothelial growth factor (VEGF) (Extended Data Fig. 2a–e), a key stimulant of angiogenesis *in vivo*⁹. There was a similar inhibitory effect of a spider toxin blocker of *Piezo1* channels, GsMTx4 (ref. 10), and a non-specific small-molecule blocker, ruthenium red⁶ (Extended Data Fig. 2f). Consistent with a relationship to endothelial cell migration, HUVEC tube formations *in vitro* and *in vivo* were suppressed by *Piezo1* depletion (Extended Data Fig. 2g–j). We therefore generated mice with a disrupted endogenous *Piezo1* gene (Extended Data Fig. 3a, b). *Piezo1*^{+/-} progeny appeared normal but *Piezo1*^{-/-} was embryonic lethal (Fig. 1a). Of 49 *Piezo1*^{-/-} embryos, the longest survival time was until embryonic day 16.5 (E16.5) and most reached only E9.5–11.5, which is a critical time for vascular development⁹. At E10.5, growth retardation was commonly observed (Extended Data Fig. 3c). At E9.5, embryos were often normal in size but yolk sac vasculature was less prominent (Fig. 1b). Although endothelial cells were present in yolk sacs of *Piezo1*^{+/-} and *Piezo1*^{-/-} embryos, the structures were different with greater disorganization and fewer defined large vessels in *Piezo1*^{-/-} (Fig. 1c). Similar observations were made at E10.5 (Fig. 1d). The data suggest *Piezo1* as a protein of critical importance in the control of vascular architecture and embryonic development.

Vascular phenotype in mice with global disruption of *Piezo1* could arise indirectly because of a requirement for *Piezo1* in non-endothelial

cells. Therefore we generated endothelial-specific disruption of *Piezo1* using Cre recombinase expressed under the *Tie2* (also known as *Tek*) promoter (Extended Data Fig. 3d–f). This disruption of *Piezo1* also caused retardation of growth (Extended Data Fig. 3g) and prevented development of normal yolk sac vasculature (Fig. 1e, f) without stopping the heart beat (Supplementary Videos 1 and 2). The data suggest there is a requirement specifically for endothelial *Piezo1*.

We considered if a mechanical stimulus relevant to endothelial biology affects *Piezo1* channels and speculated about shear stress, a frictional force arising from fluid flow. The force is sensed by endothelial cells to enable vascular development and maintain an efficient and healthy vasculature^{2,4,5,11,12}. Earlier studies have revealed multiple participating proteins and suggested sensing via a Ca²⁺-permeable non-selective cation channel^{1–3,13–19}, but the nature of the sensor itself and the molecular basis of the channel have remained controversial and elusive^{1,3}.

Piezo1 depletion and GsMTx4 treatment were found to suppress shear-stress-evoked Ca²⁺ entry in HUVECs (Extended Data Fig. 4a–f). Hepatic endothelial cells from patients undergoing surgical liver resection were also investigated and had a similar dependency on *Piezo1* (Extended Data Fig. 4g). Moreover, *Piezo1*^{-/-} embryonic endothelial cells had less shear-stress-evoked Ca²⁺ entry (Fig. 2a and Extended Data Fig. 4h, i). Furthermore, ionic current reversibly induced by shear stress had a current-voltage relationship (I–V) that was linear and which reversed near 0 mV, as expected for *Piezo1* channels⁶, and *Piezo1* depletion suppressed the current (Fig. 2b–d). In cell-attached membrane patches, negative pressure used to deliver physical force evoked unitary single channel events within less than 1 s. The unitary conductance of these channels was 25.2 ± 1.7 pS, consistent with *Piezo1* channels⁶, and *Piezo1* depletion depleted the channels (Extended Data Fig. 5). The data suggest the importance of *Piezo1* channels in shear-stress sensing and the associated Ca²⁺ entry of endothelial cells.

To investigate if *Piezo1* is sufficient for detecting shear stress, we took advantage of human embryonic kidney (HEK) 293 cells which lack endogenous *Piezo1* channels⁶. These cells had little or no shear stress response unless exogenous wild-type (WT) human *Piezo1* was expressed (Fig. 2e). A natural *Piezo1* mutant M2225R²⁰ which exhibits a slower response to stretch in cell-attached patches²¹ poorly reconstituted shear-stress-evoked Ca²⁺ entry (Fig. 2f). The data suggest that *Piezo1* is sufficient to confer shear-stress-evoked Ca²⁺ entry.

To shed light on the functional significance of shear-stress-activated *Piezo1* channels, we first tracked the subcellular localization of *Piezo1* tagged with green fluorescent protein (GFP). In static conditions it was broadly distributed but in response to shear stress there was accumulation at the leading apical lamellipodia (Extended Data Fig. 6a, b). Such apical processes²² are characteristic of early-stage alignment of endothelial cells in the direction of shear stress, a process occurring physiologically

¹School of Medicine and Multidisciplinary Cardiovascular Research Centre, University of Leeds, Leeds LS2 9JT, UK. ²School of Pharmacy, Aichi-Gakuin University, 1-100 Kusumoto, Chikusa, Nagoya 464-8650, Japan. ³Department of Hepatobiliary and Transplant Surgery, St. James's University Hospital, Leeds, LS9 7TF, UK. ⁴Cardiovascular Science, University of Sheffield, Sheffield S10 2RX, UK. ⁵Dundee Cell Products Ltd, James Lindsay Place, Dundee DD1 5JJ, UK.

*These authors contributed equally to this work.

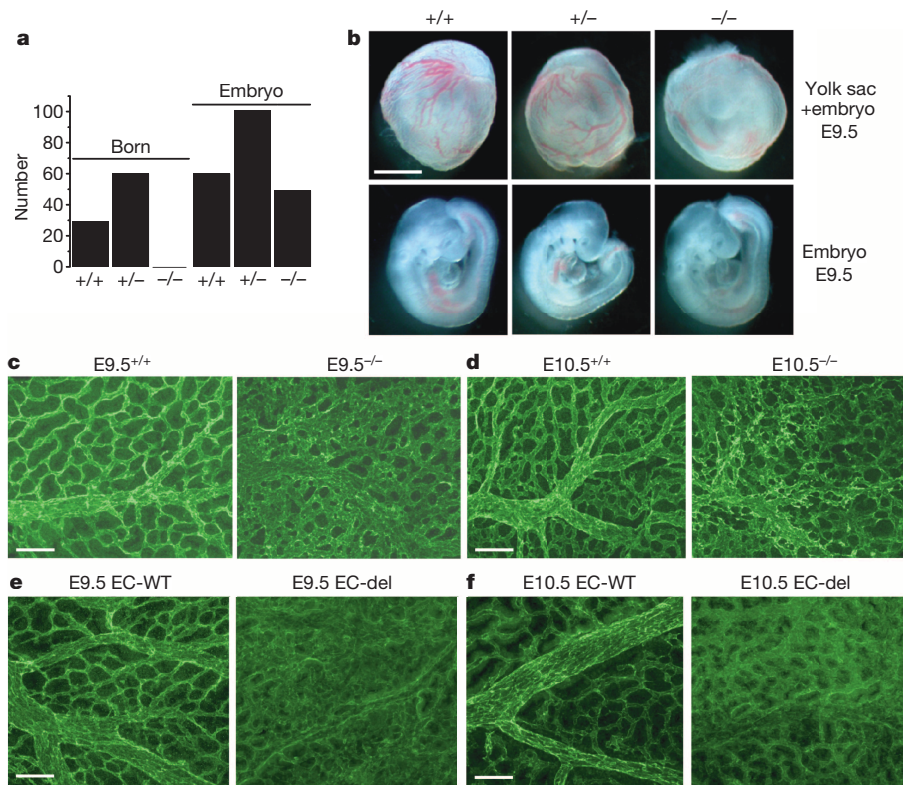


Figure 1 | Piezo1 function in murine embryos. **a**, Numbers born or detected as embryos with *Piezo1*^{+/+}, *Piezo1*^{+/-} or *Piezo1*^{-/-} genotypes. **b**, Images of sibling yolk sacs (containing embryos) and embryos at E9.5. Scale bar, 1 mm. **c**, **d**, Images of dissected *Piezo1*^{+/+} and *Piezo1*^{-/-} yolk sacs stained for CD31. **c**, E9.5 embryos. Typical of *n* = 11 (*Piezo1*^{+/+}) and *n* = 8 (*Piezo1*^{-/-}). **d**, E10.5 embryos. Typical of *n* = 9 (*Piezo1*^{+/+}) and *n* = 11 (*Piezo1*^{-/-}). **e**, **f**, As for **c**, **d**

except wild-type (EC-WT) and endothelial-specific *Piezo1*-modified (EC-del) embryos. EC-del was endothelial cell (EC)-specific deletion (del) of part of the *Piezo1* gene using Cre recombinase expressed under the *Tie2* promoter. **e**, E9.5 embryos. Typical of *n* = 6 (EC-WT) and *n* = 9 (EC-del). **f**, E10.5 embryos. Typical of *n* = 5 (EC-WT) and *n* = 11 (EC-del.). Scale bars, 100 μ m.

in blood vessels^{2,23}. We therefore measured the alignment of HUVECs in the direction of shear stress and found it to be suppressed by *Piezo1* depletion or GsMTx4 treatment (Extended Data Fig. 6c–e). Endothelial cells isolated from *Piezo1*^{-/-} embryos also showed less alignment

(Fig. 3a, b). To investigate alignment in the adult animal we took advantage of haploinsufficiency in *Piezo1*^{+/-} mice (Fig. 3c). Imaging of CD31-positive cells in arteries of these mice at 6–8 weeks revealed a striking difference in the organization of the endothelial cells: a directionless

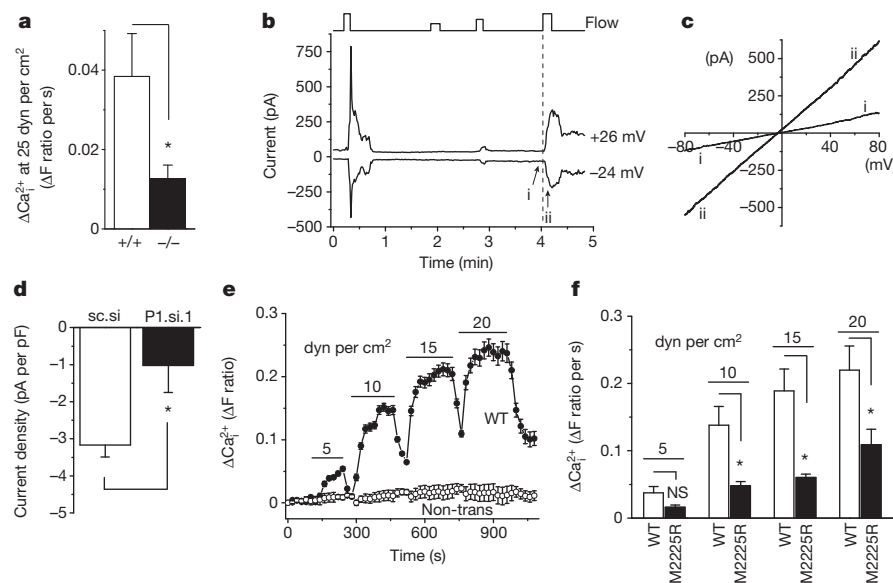


Figure 2 | Piezo1 in shear stress sensing. **a**, Ca^{2+} elevation evoked in endothelial cells from *Piezo1*^{+/+} (*n* = 6) and *Piezo1*^{-/-} (*n* = 5) E9.5 embryos. **b**, Whole-cell current in a HUVEC exposed to 12, 5, 8, and 12 μ l per s super-fusion (flow). **c**, I–V for i and ii as indicated in **b**. **d**, Paired comparison of 12 μ l per s responses at -24 mV as in **b**. Scrambled control siRNA (sc.si) *n* = 6,

Piezo1 siRNA 1 (P1.si.1) *n* = 7. **e**, Ca^{2+} in HEK 293 cells without transfection (non-trans, 5 cells) or transfected with wild-type (WT) *Piezo1*-GFP (7 cells) (*n* = 1). **f**, Mean data of the type in **e** and for the M2225R mutant (*n* = 6 each). Error bars are s.e.m.

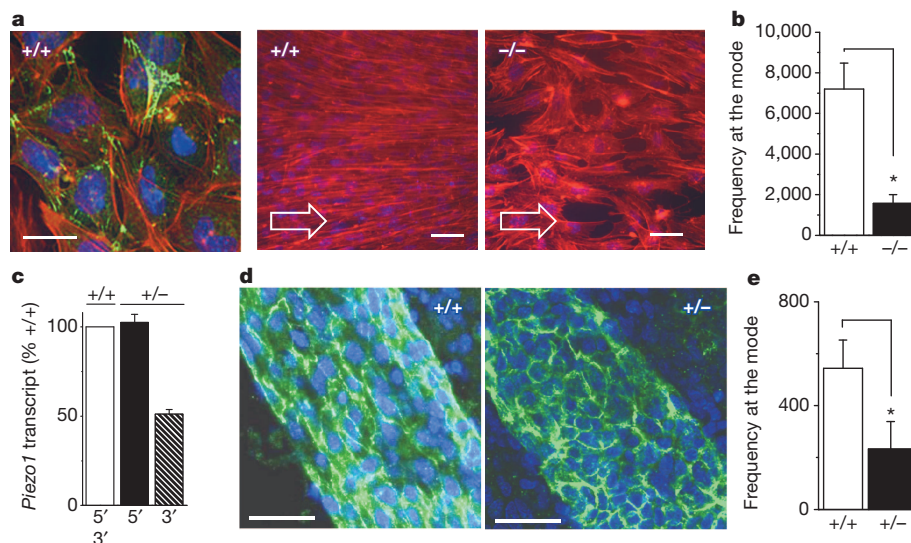


Figure 3 | The role of Piezo1 in endothelial cell alignment. **a**, Endothelial cells from E9.5 *Piezo1*^{+/+} and *Piezo1*^{-/-} embryos with no shear stress (left) or 15 dyn per cm² as indicated by arrows. CD31 (green, left only), F-actin (red), nuclei (blue). Scale bars, 50 μm. **b**, Mean data for experiments as in **a** ($n = 4$ for *Piezo1*^{+/+}, $n = 5$ for *Piezo1*^{-/-}). **c**, Cerebral artery *Piezo1* mRNA abundance

detected 5' and 3' of *Piezo1* disruption ($n = 2$ each for *Piezo1*^{+/+} and *Piezo1*^{-/-}). **d**, Cerebral arteries labelled for CD31 (green) and nuclei (blue). Scale bar, 40 μm. **e**, Quantification of *in vivo* CD31 orientation as shown in **d** ($n = 4$ each). Error bars are s.e.m.

cobblestone-like appearance in *Piezo1*^{+/-} and linear appearance in the direction of flow in *Piezo1*^{+/+} litter-mate controls (Fig. 3d, e). The data suggest Piezo1 channels as shear-stress sensors that promote endothelial cell organization and alignment in the direction of flow.

We next sought insight into the downstream mechanisms taking into account that Piezo1 channel activity was stimulated by shear stress but also important for endothelial cell migration in the absence of shear stress. In nine membrane-patch recordings we had observed occasional 25-pS channel openings in the absence of mechanical strain, consistent with low-frequency Piezo1 channel activity without exogenous force. Therefore unbiased insight into downstream pathways was sought through titanium dioxide-trapping coupled with mass spectrometry to identify regulated proteins affected by Piezo1 depletion in static and shear-stress conditions. Linked to Piezo1 under both conditions was endothelial nitric oxide synthase (eNOS, also known as NOS3) (Supplementary Table 1), a protein with important roles in vascular biology²⁴. Follow-up experiments confirmed a reduction in total eNOS protein, but more strikingly revealed abolition of VEGF-evoked phosphorylation of eNOS at serine 1177, a key enhancer of eNOS activity²⁴, in static HUVECs depleted of Piezo1 and *Piezo1*^{+/-} aorta in the absence of flow (Extended Data Fig. 7a–e). Consistent with functional relevance of coupling to eNOS, endothelial cell migration was similarly suppressed by Piezo1 depletion, eNOS depletion and NOS inhibition (Extended Data Figs 2e and

7f, g). The data suggest that in the absence of shear stress, Piezo1 activity drives endothelial cell migration through eNOS (Extended Data Fig. 7h).

Nitric oxide and eNOS played no role in endothelial cell alignment to shear stress (Extended Data Fig. 8a), but an *in silico* pathway analysis of proteomic data from endothelial cells under shear stress highlighted clusters of proteins from actin cytoskeleton (14 proteins, $P = 0.018$) and focal adhesions (16 proteins, $P = 0.002$). Relevance of these proteins was also indicated by denser actin structures in Piezo1-depleted cells (Extended Data Fig. 6c) and accumulation of Piezo1–GFP at the leading apical lamellipodia (Extended Data Fig. 6a, b) where focal adhesion turnover becomes important as the endothelial cell adjusts to achieve alignment²². Moreover, detailed inspection of HUVEC and embryo proteomic data showed significant effects on calpain-2 and many of its known substrates (Extended Data Fig. 8b and Supplementary Table 2) which are important in the structure of the actin cytoskeleton and focal adhesions²⁵. We therefore speculated that a calpain-2 system is co-regulated with Piezo1 because this system is integrated as a downstream mechanism of Piezo1. The hypothesis is consistent with calpain-2 as a Ca²⁺-activated proteolytic enzyme, previous association of calpain with Piezo1²⁶ and suggested roles of calpains in focal adhesion turnover²⁵ and endothelial cell alignment to shear stress²⁷. Moreover, disruption of a critical regulatory protein of calpain-2 (calpain small subunit 1) disturbs vascular development in the yolk sac at E10.5 (ref. 28). Consistent with this calpain

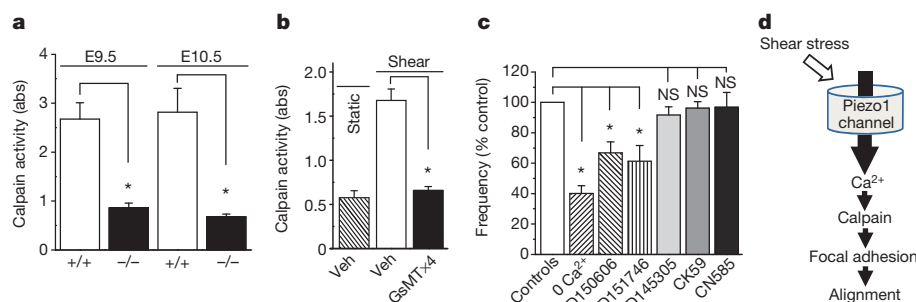


Figure 4 | Piezo1 coupling to calpain. **a**, Calpain activity indicated as absorbance (abs) in embryos at E9.5 ($n = 3$ *Piezo1*^{+/+}, $n = 3$ *Piezo1*^{-/-}) and E10.5 ($n = 3$ *Piezo1*^{+/+}, $n = 3$ *Piezo1*^{-/-}). **b**, Calpain activity in HUVECs without or with shear stress (orbital shaker) for 15 min ($n = 3$). GsMTx4 treatment, 2.5 μM. **c**, HUVEC alignment analysis as in Extended Data Fig. 6d, e.

Test conditions: nominally Ca²⁺-free Krebs solution (0 mM Ca²⁺) ($n = 3$); 3 μM PD150606 (calpain inhibitor) ($n = 3$); 20 μM PD151746 (calpain inhibitor) ($n = 3$); 20 μM PD145305 (negative control) ($n = 3$); 25 μM CK59 (CaMKII inhibitor) ($n = 3$) and 6 μM CN585 (calcineurin inhibitor) ($n = 3$). **d**, Data interpretation. Error bars are s.e.m.

hypothesis, calpain activity was significantly less in *Piezo1*^{-/-} compared with *Piezo1*^{+/-} embryos (Fig. 4a), the increase in calpain activity in response to shear stress was abolished by GsMTx4 treatment (Fig. 4b), and *Piezo1*-GFP localized to dissolving focal adhesions at the trailing edge of the cell as shear stress was applied (Extended Data Fig. 8c). The importance of *Piezo1*-mediated Ca²⁺-entry and downstream Ca²⁺-activation of calpain-2 was further indicated by the sensitivity of alignment to the absence of extracellular Ca²⁺ and presence of calpain inhibitors (Fig. 4c). Inhibitors of two other Ca²⁺-activated proteins, calcineurin and Ca²⁺/calmodulin-dependent protein kinase II (CaMKII), had no effect (Fig. 4c). The data suggest importance of calpain activation in coupling shear-stress-enhanced Ca²⁺ entry through *Piezo1* channels to endothelial cell organization and alignment via proteolytic cleavage of actin cytoskeletal and focal adhesion proteins (Fig. 4d).

Our findings have important implications for understanding vascular physiology and potentially also prevalent disease processes such as atherosclerosis and cancer in which profound alterations in shear stress and other mechanical forces are common^{2,12,29,30}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 November 2013; accepted 23 July 2014.

Published online 10 August; corrected online 12 November 2014 (see full-text HTML version for details).

- Conway, D. & Schwartz, M. A. Lessons from the endothelial junctional mechanosensory complex. *F1000 Biol. Rep.* **4**, 1 (2012).
- Chiu, J. J. & Chien, S. Effects of disturbed flow on vascular endothelium: pathophysiological basis and clinical perspectives. *Physiol. Rev.* **91**, 327–387 (2011).
- Ando, J. & Yamamoto, K. Flow detection and calcium signaling in vascular endothelial cells. *Cardiovasc. Res.* **99**, 260–268 (2013).
- Mammoto, T. & Ingber, D. E. Mechanical control of tissue and organ development. *Development* **137**, 1407–1420 (2010).
- Lucitti, J. L. *et al.* Vascular remodeling of the mouse yolk sac requires hemodynamic force. *Development* **134**, 3317–3326 (2007).
- Coste, B. *et al.* *Piezo1* and *Piezo2* are essential components of distinct mechanically activated cation channels. *Science* **330**, 55–60 (2010).
- Coste, B. *et al.* *Piezo* proteins are pore-forming subunits of mechanically activated channels. *Nature* **483**, 176–181 (2012).
- Kim, S. E., Coste, B., Chadha, A., Cook, B. & Patapoutian, A. The role of *Drosophila* *Piezo* in mechanical nociception. *Nature* **483**, 209–212 (2012).
- Olsson, A. K., Dimberg, A., Kreuger, J. & Claesson-Welsh, L. VEGF receptor signalling - in control of vascular function. *Nature Rev. Mol. Cell Biol.* **7**, 359–371 (2006).
- Bae, C., Sachs, F. & Gottlieb, P. A. The mechanosensitive ion channel *Piezo1* is inhibited by the peptide GsMTx4. *Biochemistry* **50**, 6295–6300 (2011).
- Song, J. W. & Munn, L. L. Fluid forces control endothelial sprouting. *Proc. Natl Acad. Sci. USA* **108**, 15342–15347 (2011).
- Dolan, J. M., Kolega, J. & Meng, H. High wall shear stress and spatial gradients in vascular pathology: a review. *Ann. Biomed. Eng.* **41**, 1411–1427 (2013).
- Johnson, B. D., Mather, K. J. & Wallace, J. P. Mechanotransduction of shear in the endothelium: basic studies and clinical implications. *Vasc. Med.* **16**, 365–377 (2011).
- Matthews, B. D., Overby, D. R., Mannix, R. & Ingber, D. E. Cellular adaptation to mechanical stress: role of integrins, Rho, cytoskeletal tension and mechanosensitive ion channels. *J. Cell Sci.* **119**, 508–518 (2006).
- Tzima, E. *et al.* A mechanosensory complex that mediates the endothelial cell response to fluid shear stress. *Nature* **437**, 426–431 (2005).
- Brakemeier, S., Eichler, I., Hopp, H., Kohler, R. & Hoyer, J. Up-regulation of endothelial stretch-activated cation channels by fluid shear stress. *Cardiovasc. Res.* **53**, 209–218 (2002).
- AbouAlaiwi, W. A. *et al.* Ciliary polycystin-2 is a mechanosensitive calcium channel involved in nitric oxide signaling cascades. *Circ. Res.* **104**, 860–869 (2009).
- Hartmannsgruber, V. *et al.* Arterial response to shear stress critically depends on endothelial TRPV4 expression. *PLoS ONE* **2**, e827 (2007).
- Yamamoto, K. *et al.* Impaired flow-dependent control of vascular tone and remodeling in P2X4-deficient mice. *Nature Med.* **12**, 133–137 (2006).
- Zarychanski, R. *et al.* Mutations in the mechanotransduction protein *PIEZO1* are associated with hereditary xerocytosis. *Blood* **120**, 1908–1915 (2012).
- Bae, C., Gnanasambandam, R., Nicolai, C., Sachs, F. & Gottlieb, P. A. Xerocytosis is caused by mutations that alter the kinetics of the mechanosensitive channel *PIEZO1*. *Proc. Natl Acad. Sci. USA* **110**, E1162–E1168 (2013).
- Li, S. *et al.* The role of the dynamics of focal adhesion kinase in the mechanotaxis of endothelial cells. *Proc. Natl Acad. Sci. USA* **99**, 3546–3551 (2002).
- Langille, B. L. & Adamson, S. L. Relationship between blood flow direction and endothelial cell orientation at arterial branch sites in rabbits and mice. *Circ. Res.* **48**, 481–488 (1981).
- Balligand, J. L., Feron, O. & Dessy, C. eNOS activation by physical forces: from short-term regulation of contraction to chronic remodeling of cardiovascular tissues. *Physiol. Rev.* **89**, 481–534 (2009).
- Lebart, M. C. & Benyamin, Y. Calpain involvement in the remodeling of cytoskeletal anchorage complexes. *FEBS J.* **273**, 3415–3426 (2006).
- McHugh, B. J. *et al.* Integrin activation by Fam38A uses a novel mechanism of R-Ras targeting to the endoplasmic reticulum. *J. Cell Sci.* **123**, 51–61 (2010).
- Miyazaki, T., Honda, K. & Ohata, H. Requirement of Ca²⁺ influx- and phosphatidylinositol 3-kinase-mediated m-calpain activity for shear stress-induced endothelial cell polarity. *Am. J. Physiol. Cell Physiol.* **293**, C1216–C1225 (2007).
- Arthur, J. S., Elce, J. S., Hegadorn, C., Williams, K. & Greer, P. A. Disruption of the murine calpain small subunit gene, *Capn4*: calpain is essential for embryonic development but not for cell growth and division. *Mol. Cell. Biol.* **20**, 4474–4481 (2000).
- Zhuang, X., Cross, D., Heath, V. L. & Bicknell, R. Shear stress, tip cells and regulators of endothelial migration. *Biochem. Soc. Trans.* **39**, 1571–1575 (2011).
- Hahn, C. & Schwartz, M. A. Mechanotransduction in vascular physiology and atherogenesis. *Nature Rev. Mol. Cell Biol.* **10**, 53–62 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements The research was supported by research grants from the Wellcome Trust, the Medical Research Council, the Leeds Teaching Hospitals Trust Charitable Foundation, and the British Heart Foundation. B.H. was supported by a Scholarship from the University of Leeds and the China Scholarship Council. A.J.H. was supported by a BBSRC PhD Studentship. R.S.Y. was supported by a Cancer Research UK Clinical Fellowship. L.A.W. was supported by a BBSRC-AstraZeneca PhD Studentship. M.A.B. was supported by a British Heart Foundation Fellowship.

Author Contributions J.L. initiated the experimental studies of *Piezo1* and was the primary contributor to experiments on endothelial cell tube formation, *Piezo1* gene-modified mice and *Piezo1* overexpression. B.H. initiated the experimental studies of shear stress and was the primary contributor to experiments on shear-stress-evoked Ca²⁺ responses, *Piezo1* redistribution and *Piezo1*-dependence of endothelial cell alignment. J.L. and B.H. addressed the calpain hypothesis. S.T. initiated the proteomic experiments and nitric oxide synthase studies. K.M. performed patch-clamp experiments. A.S., R.S.Y., N.Y.Y., L.M.K., Y.M., L.A.W., B.R., A.B., M.J.L., A.J.H., D.A.L.C., J.B., P.A. and R.M.C. also contributed to experiments or prepared cells, mice or reagents. J.L., B.H., S.T., K.M., H.L., Z.F. and A.J.H. analysed data, interpreted data and developed methods. K.R.P. provided essential material. J.L., B.H., S.T., K.M., H.R.K., M.T.K., M.A.B., T.N.D., P.C.E. and J.F.X.A. provided intellectual input. All authors commented on the manuscript. D.J.B. initiated the project, generated research funds and ideas, led and coordinated the project, interpreted data and wrote the paper.

Author Information The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (dataset identifier: PXD001099 and DOI 10.6019/PXD001099). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.J.B. (d.j.beech@leeds.ac.uk).

METHODS

Piezo1-modified mice. All animal use was authorized by both the University of Leeds Animal Ethics Committee and by The Home Office, UK. Project licenses used in this work were 40/3557 and 40/2946. All animals were maintained in Optimize individually ventilated cages (Animal Care Systems) at 21°C 50–70% humidity, light/dark cycle 12 h/12 h on RM1 diet (Special Diet Services, Witham, UK) *ad libitum* and bedding of Pure-o Cell (Datesand, Manchester, UK) with enrichment of Sizzlenest (International Product Supplies, London, UK). Piezo1 Knockout First (with conditional potential) embryonic stem cells (Piezo^{tm1(KOMP)Wtsi, clone ID EPD0500_5_F12}) were obtained from the KOMP Repository (<http://www.komp.org>) and injected into C57BL/6J blastocysts. The injected blastocysts were transplanted into pseudopregnant CD-1 recipient females. Chimaeric offspring were crossed to C57BL/6J mice to obtain germline transmission. Offspring were a minimum of N₃ on a C57BL/6J background before being intercrossed to generate *Piezo1* homozygotes. To generate endothelial-specific disruption of *Piezo1* we first crossed mice carrying global disruption of *Piezo1* with mice expressing FLP1 recombinase (B6.SJL-Tg(ACFPLPe)9205Dym/J mice from the Jackson Laboratory) to delete the *lacZ* insertion flanked by FRT sites (Extended Data Fig. 3a). Deletion of *lacZ* was confirmed (Extended Data Fig. 3d) and FLP1 recombinase gene was bred out. The mice were crossed with mice expressing Cre recombinase under the *Tie2* (*Tek*) promoter (B6.Cg-Tg(Tek-cre)12Flv/J from the Jackson Laboratory). Males carrying *Tie2*-Cre were used to transmit *Tie2*-Cre to progeny and obtain endothelial deletion in embryos. Successful Cre-driven deletion and depletion of *Piezo1* mRNA was confirmed by PCR (Extended Data Fig. 3e, f).

Genotyping and tissue isolation. Germline transmission and all genotyping were determined by PCR analysis of DNA in ear-notches (see Extended Data Fig. 3 for example data and Supplementary Table 3 for primer sequences). Embryos and yolk sacs were harvested from timed pregnancies at the indicated day post-coitus; embryonic (E) day 0.5 was the time when a vaginal plug was observed. Thoracic aorta was obtained from 6–8-week-old male mice after killing by CO₂ asphyxiation and cervical dislocation in accordance with Schedule 1 Code of Practice, UK Animals Scientific Procedures Act 1986.

Ultrasound. Mice were anaesthetized with isoflurane and each embryo was observed using a Vevo 2100 Imaging System (VisualSonics). The anatomical position of each embryo was recorded. After euthanasia by Schedule 1 procedure, embryos were dissected and genotyped.

Mouse embryonic endothelial cells. E9.5 embryos were individually transferred to an eppendorf tube containing 1 ml of a solution containing collagenase (1.5 mg ml⁻¹) and DNase (25 µg ml⁻¹) in Hanks balanced salt solution and incubated at 37°C for 60 min with occasional gentle agitation by pipetting. The obtained cell suspension was pelleted by centrifugation for 3–5 min at 3,000 r.p.m. and the cell pellet washed with 1 ml PBS. Cells were seeded in Ibidi slides with EGM-2 medium containing 5% serum and used for experiments after 24 h. Cells staining positively for CD31 (Fig. 3a) or responding to VEGF (Extended Data Fig. 4h, i) were classed as endothelial cells and used for analysis.

Fresh mouse artery and cultured mouse lung microvascular endothelial cells. Thoracic aorta was obtained from 6–8-week-old male C57BL/6J mice after killing by CO₂ asphyxiation and cervical dislocation in accordance with Schedule 1 Code of Practice, UK Animals Scientific Procedures Act 1986. The mice were also used for endothelial cell isolation from lungs by immunoselection with CD146 antibody-coated magnetic beads (Miltenyi Biotec, UK). Lungs were harvested in ice-cold Hanks' balanced salt solution (HBSS), finely minced, and digested in HBSS containing 0.18 units per ml collagenase (10 mg ml⁻¹; Worthington, USA) for 45 min at 37°C. The digested tissue was filtered through a 70-µm cell strainer and centrifuged at 1,000 r.p.m. for 10 min. The cell pellet was washed with PBS/0.05% BSA, centrifuged, resuspended in 90 µl PBS/0.05% BSA, and incubated with 15 µl CD146 antibody-coated beads at 4°C for 20 min. Bead-bound cells were separated from non-bead-bound cells using a magnet. Bead-bound cells (CD146 positive) were resuspended in 2 ml Endothelial Cell Medium MV2 (PromoCell), with the manufacturer's supplement, gentamicin, amphotericin-B, and 10% FCS, and seeded onto fibronectin coated plates with a full media change at 2 and 24 h post-isolation. The cell population tested positive for the endothelial markers eNOS, Tie2 and CD102.

Human endothelial cells. Human umbilical vein endothelial cells (HUVECs) were validated by positive staining with anti-CD31 antibody, VEGF responsiveness and alignment to shear stress. Human nucleotide sequences were detected in the cells, confirming their human origin. No mycoplasma infection was detected. The HUVECs were cultured in EGM-2 growth medium supplemented with EGM-2 bullet kit (Lonza). Cells were maintained at 37°C in a humidified atmosphere containing 5% CO₂. For human liver endothelial cells, participants undergoing liver resection for colorectal metastases at St James's University Hospital, Leeds Teaching Hospitals NHS Trust, provided fully-informed written consent. The work was carried out under approval granted by the local ethics committee (Ref 10/H1306/82) and the study was adopted into the United Kingdom Clinical Research Network portfolio (ID 9442).

Tissue was collected immediately following resection. A section of normal liver tissue was immediately taken from a macroscopically normal area of the surgical specimen, > 2.5 cm from any tumours, and with subsequent microscopic confirmation. The samples were stored and transported in cold EGM-2 growth medium supplemented with EGM-2 bullet kit (Lonza) on ice. Endothelial cells were prepared using a protocol adopted from a previous study³¹. Briefly, 1 g of tissue was weighed and minced in a Petri dish. The tissue was resuspended in 9 ml 0.1% collagenase and 1 ml 2.5 U per ml dispase solution and incubated for 45 min at 37°C in a water bath under continuous agitation. At the end of enzymatic digestion the sample was passed through 100-µm then 40-µm cell strainers to remove major debris and then washed twice in MACS buffer consisting of phosphate-buffered saline (PBS), EDTA 2 mM, and 0.1% serum. The pellet was resuspended in 20 ml of red blood cell lysis buffer (TRIS base 17 mM, NH₄Cl 140 mM) for 10 min at room temperature. Following a final wash in MACS buffer the cells were incubated with 100 µl of dead cell removal paramagnetic microbeads per 1 × 10⁷ cells (Miltenyi Biotec) in buffer at room temperature for 15 min. The suspension was then passed through a MACS cell separation column in a magnetic field to retain apoptotic cells, dead cells and debris. The live cell fraction was eluted and incubated with 20 µl FcR blocking reagent and 20 µl anti-CD31-conjugated paramagnetic microbeads per 1 × 10⁷ cells at 4°C for 15 min. This suspension was then passed through a column in a magnetic field and the unbound cells eluted with buffer. Once removed from the magnetic field, the column was washed with buffer to elute CD31-positive cells. The approach was validated by flow cytometry and immunostaining. Cells were incubated at 37°C in 5% CO₂ and used up to passage 3.

Human dermal (juvenile foreskin), bladder, cardiac, colonic, and pulmonary microvascular and pulmonary artery and umbilical artery endothelial cells were from PromoCell. Cells were maintained at 37°C in a humidified atmosphere containing 5% CO₂. For late outgrowth endothelial progenitor cells (LEPCs) venous blood samples were drawn from healthy volunteers into EDTA-coated tubes and processed immediately. Ethical approval and consent were obtained. Blood was mixed with an equal volume of PBS and layered onto Ficoll Paque PLUS (GE Healthcare) before density gradient centrifugation. Peripheral blood mononuclear cells were aspirated from the buffy layer and washed twice with PBS. Mononuclear cells (5 × 10⁶ per ml) were then suspended in EGM-2 growth medium supplemented with EGM-2 bullet kit (Lonza) with 10% fetal calf serum (FCS) and 2 ml of cell suspension was added to a fibronectin-coated 6-well plate (Becton Dickinson). The cells were incubated at 37°C in 5% CO₂ and culture medium was changed every day for the first week, then on alternate days. After 2–3 weeks, colonies of LEPCs were observed; on day 28, cells were detached using trypsin/EDTA 0.025% solution (Gibco BRL) and placed into large fibronectin coated vessels. The cells exhibited contact inhibition and were capable of serial passage. Flow cytometry revealed that 89% of cells co-expressed vascular endothelial growth factor (VEGF) receptor-2 and CD31.

Cell migration and tube formation assays. Migration assays were performed using a modified Boyden chamber (8-µm pores; BD Biosciences, UK). HUVECs transfected with siRNA were serum-starved in EGM-2 for 4 h, resuspended at 5 × 10⁴ cells per ml in EGM-2 containing 0.4% FCS and loaded in the upper chamber. The lower chamber contained 0.4% FCS supplemented with the chemoattractant 20 ng ml⁻¹ VEGF (Sigma). After incubation for 4 h at 37°C in a 5% CO₂ incubator, the inserts were fixed with 70% ethanol for 2 min. Cells on duplicate membranes were scraped from the upper surface and migrating cells underneath were stained with haematoxylin and eosin and evaluated by counting cells in 6 randomly chosen fields under light microscopy. For fibroblast-endothelial cell co-culture tube-formation assays, normal human dermal fibroblasts (NHDFs) were plated at 2 × 10⁵ cells per well with fibroblast growth medium (FGM) in 96-well plates (Greiner bio-one) and incubated at 37°C in 5% CO₂. The medium was removed from NHDFs on day 4 and transfected HUVECs were loaded on top at 5.4 × 10⁴ cells per well in endothelial cell growth medium supplemented with 2% serum and 3 ng ml⁻¹ VEGF. Co-cultures were incubated for a further 7 days at 37°C in 5% CO₂. Tubes were analysed following fixing and staining of endothelial cells with anti-CD31 antibody. Quantification of tube formation was conducted by using IncuCyte angiogenesis v2.0 image analysis (Essen Bioscience). For *in vivo* tube-formation studies, ice-cold Matrigel (0.5 ml per plug; BD Biosciences) was mixed with VEGF (100 ng ml⁻¹), EGM-2 and 5 × 10⁵ HUVECs transfected with P1.si.1 or sc.si and then subcutaneously injected into the flanks of 6–8-week-old male immunodeficient CD1 nude mice (Charles River Labs, Boston, MA). After 7 days, mice were anaesthetized and Matrigel plugs were carefully dissected away from the surrounding adherent tissue, washed with PBS and fixed in 4% paraformaldehyde, embedded in paraffin, sectioned, and stained with haematoxylin and eosin. Digital images were obtained by bright-field light microscopy. Blood vessels in the plug were quantified by ImageJ software (NIH). All animal experiments were performed in accordance with ethical approval under a UK Home Office licence.

DNA constructs and transfection. Human Piezo1 IRES GFP (from J Wood) was used as a PCR template to create Piezo1-GFP using inverse primers (forward:

5'GAGGGTGGAGGTGGAACAACCATGGTGAGCAAGGGCGGCC; reverse: 5'TC CACCTCCACCTCTCTCAGAGTCCAC) and pHUSION (NEB, Herts, UK) to delete the IRES cDNA and insert a linker of four glycine residues. The PCR product was recombined using Clontech In-Fusion HD cloning kit (Takara Bio Europe, France). A Kozak sequence was inserted into the Piezo-GFP 5' region using mutagenic primers (forward: 5'TCCACCATGGAGCCGACGTCGT; reverse 5'AG CACGTGCGGCTCCATGGTGGGA) and pHUSION. M2225R Piezo1-GFP mutagenesis was performed using mutagenic primers (M2225R Forward: ATGAGCC GCTGTTCACAGGAGCGCCC; M2225R Reverse: GGGCGCTCTGGTGAA CAGCGGCTCAT) and pHUSION. All clones were sequenced to confirm accuracy and identity. HUVECs and HEK 293 cells (GripTite 293 cells) were transfected with FuGENE HD (Promega) using 1 and 0.3 µg of plasmid and measurements were made after 48 and 72 h respectively.

RNA interference and RT-PCR. Endothelial cells at 90% confluence were transfected with 20 nM siRNA using Lipofectamine 2000 in OptiMEM as per the manufacturer's instructions (Invitrogen). Sequences of siRNA probes are given in Supplementary Table 3. The control siRNA was from Ambion or Dharmacon depending on the source of the test siRNA (Extended Data Fig. 2a). Fresh EGM-2 growth medium was added after 3–4 h and the cells were analysed 48 h after transfection. To validate effectiveness of siRNA probes, mRNA was isolated and quantified by real-time RT-PCR. Total RNA was extracted using a Tri-reagent protocol followed by DNase I (Ambion) treatment. 1 µg of total RNA was used for reverse transcription (RT) based on oligo-dT primers and AMV RT enzyme. The specificity of PCR was verified by reactions without RT (–RT) and melt-curve analysis. Sequences of PCR primers are given in Supplementary Table 3. PCR products were electrophoresed on 2% agarose gels containing ethidium bromide and sequenced to confirm identity. Real-time PCR was carried out using a Lightcycler (Roche). For RNA isolation from cerebral arteries, vessels were dissected from the brain and surrounding connective tissue, snap frozen and RNA extracted as indicated above.

Proteomics. Cells were processed 2 days after transfection or embryos were snap-frozen after dissection from the animal. Samples for label-free phosphoproteomic analysis were processed by modified Filter-Aided Sample Preparation (FASP) method³². Briefly, 1.3 mg protein lysate for each sample was diluted in 8 M urea, 20 mM DTT in 100 mM Tris/HCl pH 8.5 (FASP1 buffer) and concentrated using Vivacon 500, 30k MWCO HY filter vials (Sartorius Stedim Biotech, VN01H22). After several exchanges in the FASP1 buffer, samples were diluted in 100 mM Tris/HCl pH 8.5 and reduced by 50 mM iodoacetamide. Excess iodoacetamide was removed by centrifugation and after buffer exchange into 100 mM triethyl ammonium bicarbonate, trypsin digest was performed overnight followed by second digestion for 5 h. Resulting peptide mixture was washed with 0.5 M NaCl, acidified with 10% TFA and desalted³³. Phosphopeptides in the mixture were enriched by incubation with TiO₂ beads equilibrated in 1 M glycolic acid, 80% ACN, 5% TFA, washed and eluted by 0.5% NH₄OH. Acidified eluates were dried under vacuum and resuspended in 1% FA. Peptides were separated using an Ultimate 3000 RSLC (Thermo Scientific) nanoflow LC system. Triplicates of each sample were loaded onto an Acclaim PepMap100 nanoViper C18 trap column (100-µm inner-diameter, 2 cm; Thermo Scientific) for trap enrichment, then peptides were eluted onto an Acclaim PepMap RSLC nanoViper, C18 column (75-µm, 15 cm; ThermoScientific) and subjected to a 65 min linear gradient of 2–40% solvent B (80% acetonitrile with 0.08% formic acid) with a constant flow of 300 nl min^{−1}. The HPLC system was coupled to a linear ion trap Orbitrap hybrid mass spectrometer (LTQ-Orbitrap Velos, Thermo Scientific) via a nano-electrospray ion source (Thermo Scientific), with the spray voltage set to 1.2 kV, and the temperature of the heated capillary at 250°C. Spectra acquisition was performed essentially as described previously³⁴ using the Xcalibur software (Thermo Scientific).

Label-free quantification of the acquired spectra was performed by the Progenesis LC-MS software, which uses wavelet-based filtering to smooth the peak envelopes and screen noisy areas and transforms the profile data of the MS scans to peak lists comprising *m/z* (mass-to-charge ratio) and abundance. One sample run was selected as a reference and the automatic Alignment Algorithm was applied to enable alignment of all the LC-MS runs to the reference run. Features with only one charge or more than seven charges were excluded from further analyses, while all remaining features were used for normalization. The database search was performed with MASCOT (version 2.3.2, Matrix Science, London, UK), using trypsin as the enzyme and allowing for cleavage N-terminal to proline residues and between aspartic acid and proline residues. Any peptides with conflicting assignments were resolved, either by assignment to one of identical proteins or by assignment to proteins with the largest number of peptides already present, following Occam's Razor principle. For protein quantification, all peptides of an identified protein were included and the total cumulative abundance was calculated by summing the abundance of all peptides allocated to the respective protein. Calculation of the protein *P* value by one-way ANOVA was then performed on the sum of the normalized abundance across all runs. *In silico* pathway analysis was performed by the

Database for Annotation, Visualization and Integrated Discovery³⁵ (<http://david.abcc.ncifcrf.gov/>).

Western blotting. Cells were transfected and harvested in lysis buffer containing 10 mM Tris, pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP-40, MiniComplete protease inhibitors (Roche), and PhosSTOP phosphatase inhibitors (Roche). Equal protein amounts were loaded on 8% gels and resolved by electrophoresis. Samples were transferred to PVDF membranes and labelled overnight with primary antibody: anti-β-actin (0.2 µg ml^{−1}, Santa Cruz); anti-total-eNOS (0.2 µg ml^{−1}, BD Biosciences); anti-p(S1177)-eNOS (0.2 µg ml^{−1}, BD Biosciences); or anti-Piezo1 (0.3 µg ml^{−1}, Proteintech). Horse radish peroxidase-conjugated donkey anti-mouse or anti-rabbit secondary antibody (Jackson ImmunoResearch) and SuperSignal Femto detection reagent (Pierce and Warriner) were used for visualization. Densities of protein bands were quantified by ImageJ software. Catalogue information for antibodies: anti-β-actin (0.2 µg ml^{−1}, Clone C4, sc-47778; Santa Cruz); anti-total-eNOS (0.2 µg ml^{−1}, Clone 3/eNOS/NOS Type III, Cat. #610297; BD Biosciences); anti-p(S1177)-eNOS (0.2 µg ml^{−1}, Cat. #612392; BD Biosciences); or anti-Piezo1 (0.3 µg ml^{−1}, Proteintech).

Immunofluorescence and live-cell imaging. Cells were fixed with 2% paraformaldehyde for 10 min and permeabilized with 0.1% Triton X-100 for 10 min at room temperature. Non-specific sites were blocked using 10% donkey serum in PBS for 1 h at room temperature. Cells were then incubated in 2% donkey serum in PBS containing mouse anti-human CD31/PECAM-1 (Dako, clone JC70A) at 1:500 dilution for 1 h at room temperature. After washing with PBS, cells were incubated with DyLight 649-conjugated Affinipure Donkey anti Mouse IgG (Jackson Immuno Research Laboratories) for 30 min at room temperature. Cells were mounted with Prolong Gold Antifade Reagent (Invitrogen) and visualized using a DeltaVision deconvolution system (Applied Precision Instruments, Seattle, WA) on an Olympus IX-70 inverted microscope fitted with ×60 UPLAN objective (NA 1.35). For F-actin staining, cells were incubated with 1:250 rhodamine phalloidin (Cytoskeleton Inc.) for 30 min at room temperature. Cell nuclei were labelled with DAPI (4',6-diamidino-2-phenylindole) with the excitation wavelength of 350 nm. Piezo1-GFP was observed at the excitation wavelength of 488 nm. For tracking Piezo1-GFP in live HUVECs, cells were seeded in an Ibidi µ-Slide VI^{0.4} and imaged 2 days after transfection using the DeltaVision deconvolution system and an excitation wavelength of 488 nm. Images were sampled every 15 s for 10 min before 15 dyn per cm² was applied and imaging continued for 50 min. To quantify the GFP fluorescence, its intensity in multiple squares (5 or 9 µm² depending on the shape of the apex) was measured using ImageJ software for each region of the cell indicated in Extended Data Fig. 6a. The number of squares for each region was between 3 and 5 and the average intensity for all squares in each region was used. Yolk sacs and mouse cerebral arteries were dissected in cold PBS and fixed in 4% paraformaldehyde overnight at 4°C or 1 h at room temperature and permeabilized with 0.1–0.3% Triton X-100 for 1 h, blocked in 2% BSA and 0.1% Triton X-100 for 2 h at room temperature, followed by labelling with anti-CD31/PECAM-1 antibody (ab28364 Abcam) overnight at 4°C. Yolk sacs or cerebral arteries were then incubated with DyLight 488-conjugated Affinipure Donkey anti rabbit IgG for 1 h at room temperature and mounted with Prolong Gold Antifade Reagent. Images were captured using a confocal microscope (Zeiss LSM 700).

Electrophysiology. Borosilicate glass capillaries with an outside diameter of 1 mm and an inside diameter of 0.58 mm (Harvard Apparatus, Holliston, MA, USA) were used as the basis for patch pipettes. Pipettes were pulled using a PP-830 vertical 2-stage pipette-puller (Narishige, Tokyo, Japan). Pipette resistances after fire-polishing and filling with pipette solution were 3–5 MΩ. Pipettes were mounted on a CV-4 head-stage (Molecular Devices, Sunnyvale, CA, USA) connected to a 3-way coarse manipulator and micromanipulator (Mitutoyo, Japan). The electrode was a Ag/AgCl wire. Electrical signals were amplified and recorded using an Axopatch 1D amplifier and pCLAMP 10 software (Molecular Devices). Data were filtered at 1 kHz and sampled digitally at 3 kHz via a Digidata 1322A analogue to digital converter (Molecular Devices). Analysis was performed off-line using Clampfit 10.2 (Molecular Devices) and Origin 7.5 software (OriginLab, Northampton, MA). Recordings were made at room temperature. Recordings in cell-attached patch mode used a bath solution containing (mM): 140 KCl, 10 D-glucose, 10 HEPES, 1 MgCl₂, titrated to pH 7.4 with KOH. The patch pipette solution contained (mM): 130 NaCl, 5 KCl, 10 TEA (tetraethylammonium) Cl, 8 D-glucose, 10 HEPES, 1.2 MgCl₂, 1.5 CaCl₂, titrated to pH 7.4 with NaOH. Negative pressure was applied to the inside of the patch pipette via a valve and syringe coupled to a calibrated pressure transducer. For whole-cell recording the patch pipette solution contained (mM): 95 Na-aspartate, 40 CsCl, 10 HEPES, 1 MgCl₂, 1 CaCl₂, 5 EGTA, 10 TEACl, titrated to pH 7.2 with CsOH; the bath solution contained (mM): 130 NaCl, 5 KCl, 8 D-glucose, 10 HEPES, 1.2 MgCl₂, 1.5 CaCl₂, titrated to pH 7.4 with NaOH. Correction was made for a calculated −9 mV liquid-liquid junction potential. The Cl[−] equilibrium potential was at −24 mV. The superfusion pipette had a diameter of ~350 µm.

Intracellular Ca²⁺ measurement. Cells were incubated with fura-2AM for 1 h at 37°C followed by a 0.5 h wash at room temperature (21 ± 2°C) on Ibidi µ-Slide VI^{0.4}.

Measurements were made at room temperature on a Zeiss Axiovert fluorescent microscope equipped with $\times 20$ (NA 0.75) or $\times 40$ (NA 1.3, oil) Fluor objective and excitation light from a xenon lamp selected by a monochromator (Till photonics, Germany). Emitted light was collected via an emission filter and images captured by an Orca-ER digital camera (Hamamatsu, Japan). The extracellular (superfusion) recording solution contained (mM): 130 NaCl, 5 KCl, 8 D-glucose, 10 HEPES, 1.2 MgCl_2 , 1.5 CaCl_2 , titrated to pH 7.4 with NaOH. For Extended Data Fig. 4a, b, 2 ng ml^{-1} VEGF was included. The change (Δ) in intracellular calcium (Ca^{2+}) concentration above baseline is shown by the ratio of fura-2 fluorescence (F) emission intensities for 340 and 380 nm excitation (ΔF ratio). For statistical comparisons, the total ΔF ratio above baseline was calculated for each shear stress and divided by the time period over which the ΔF ratio was measured (ΔF ratio per s).

Calpain activity. The assay detected cleavage of calpain substrate Ac-LLY-AFC (Abcam). Cells were counted, pelleted by centrifugation, and resuspended in 100 μl of extraction buffer and incubated on ice for 20 min. Gentle mixing was achieved by tapping several times during the incubation. Centrifugation was applied for 1 min in a microcentrifuge (10,000g) and the cell lysate was diluted in 85 μl of extraction buffer and transferred to wells in a 96-well plate. Embryos were homogenized before dilution in extraction buffer. All inputs were standardized according to total protein content. 10 μL of $10\times$ reaction buffer and 5 μl of calpain substrate were added to each assay well. Incubation occurred at 37°C for 1 h in the dark, after which measurements were made on a plate-reader equipped with a 400 nm excitation filter and 505 nm emission filter. Absorbance values in arbitrary units are presented after subtraction of background.

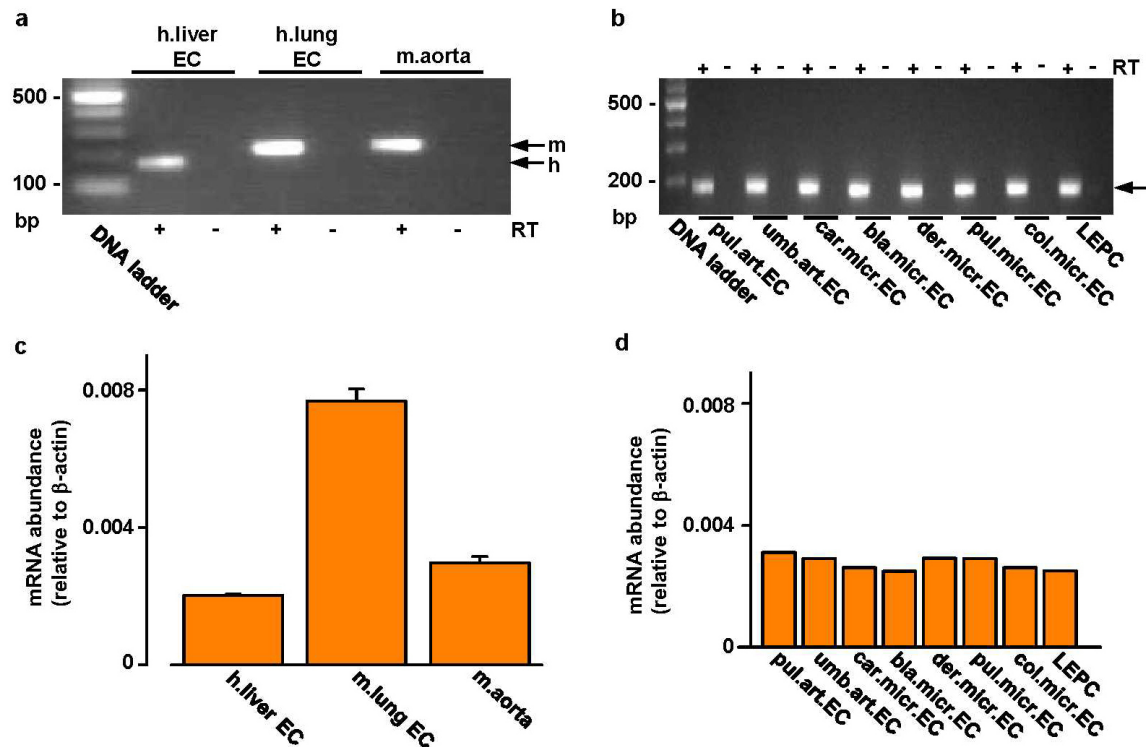
Shear stress. Shear stress was achieved in microfluidic chambers (<http://ibidi.com/>) or on an orbital shaker (210 r.p.m.) that generated swirling motion of medium around the edges of the wells producing tangential shear stress, which results in cell elongation and alignment along the edges³⁶. Shear stress is specified in dyn per cm^2 , where 1 dyn per cm^2 is 0.1 Pa or 0.1 N per m^2 . For the orbital shaker experiments, cells 2 days after transfection were replated at 70% confluency onto 6-well plates. After 4 h, cells were subjected to shear stress for up to 60 h. In experiments of Fig. 4c where extracellular Ca^{2+} was omitted, Krebs solution was used and it contained (mM): NaCl, 125; KCl, 3.8; NaHCO_3 , 25; MgSO_4 , 1.5; KH_2PO_4 , 1.2; D-glucose, 8; CaCl_2 , 0 or 1.2; EDTA, 0.02.

Reagents. GsMTx4, a toxin from the tarantula *Grammostola spatulata*, was from Peptide Institute Inc. PD145305 (3-phenyl-2-sulfanylpropanoic acid) and PD151746 (3-(5-fluoro-3-indolyl)-2-mercapto-(Z)-2-propenoic acid) were from Santa Cruz Biotechnology. PD150606 (3-(4-iodophenyl)-2-mercapto-(Z)-2-propenoic acid), CK59 (2-(2-hydroxyethylamino)-6-aminohexylcarbamate tert-butyl ester-9-isopropylpurine) and CN585 (6-(3,4-dichlorophenyl)-4-(N,N-dimethylaminoethylthio)-2-phenyl-pyrimidine) were from EMD Millipore. L-NMMA (N-methyl-L-arginine) and ruthenium red were from Sigma-Aldrich. The solvent for each stock solution was dimethylsulphoxide except in the cases of GsMTx4, ruthenium red and L-NMMA when it was water. Inhibitors were pre-incubated with cells for 1 h before stimulating cell activity.

Cell orientation analysis. Using ImageJ software, images were rotated to the direction of applied shear stress or blood flow. For analysis of intact arteries, images were analysed only within the outer arterial wall. Images were processed using a Difference of Gaussian plugin to define cell edges (<http://www.sussex.ac.uk/gdsc/intranet/microscopy/imagej/utility>). Automated quantification of cell orientation relative to the direction of shear stress was determined using OrientationJ software, also an ImageJ plugin³⁷ (<http://bigwww.epfl.ch/demo/orientation/>). OrientationJ produced a histogram of all local angles in each image. A Gaussian distribution curve was fitted to each arising histogram. The baseline-subtracted frequency maximum at the mode of the distribution was determined.

Statistical analysis. All averaged data are presented as mean \pm s.e.m. Data were produced in pairs (test and control) and the raw data pairs were compared statistically using a one-tailed *t*-test for the hypothesis that the test condition decreased the response and a two-tailed *t*-test for the hypothesis that the response changed in either direction (that is, decrease or increase). Normal distribution of data was assumed but equal variance was not. For one data set we used one-way ANOVA followed by Tukey posthoc test (Extended Data Fig. 4d). In all cases, statistically significant difference is indicated by $*P < 0.05$ and no significant difference is indicated by NS ($P > 0.05$). Statistical tests were performed using OriginPro 8.6 software. Analysis of proteomic data was an exception where ANOVA was used as indicated in the specific section for this work. The letter *n* is used to denote the number of independent biological experiments and does not denote replicates within one such experiment (the number of replicates was always ≥ 2). Randomization of samples was not applied. Blinding was applied to Ca^{2+} measurement and cell alignment studies of embryonic endothelial cells and patch-clamp studies of HUVECs after transfection with siRNA. The Source Data file contains the raw data underlying graphical presentations.

- van Beijnum, J. R., Rousch, M., Castermans, K., van der Linden, E. & Griffioen, A. W. Isolation of endothelial cells from fresh tissues. *Nature Protocols* **3**, 1085–1091 (2008).
- Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature Methods* **6**, 359–362 (2009).
- Thingholm, T. E., Jorgensen, T. J., Jensen, O. N. & Larsen, M. R. Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nature Protocols* **1**, 1929–1935 (2006).
- Babaei, F. *et al.* Novel blood collection method allows plasma proteome analysis from single zebrafish. *J. Proteome Res.* **12**, 1580–1590 (2013).
- Huang Da W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
- Warboys, C. M. *et al.* Disturbed flow promotes endothelial senescence via a p53-dependent pathway. *Arterioscler. Thromb. Vasc. Biol.* **34**, 985–995 (2014).
- Rezakhaniha, R. *et al.* Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy. *Biomech. Model. Mechanobiol.* **11**, 461–473 (2012).

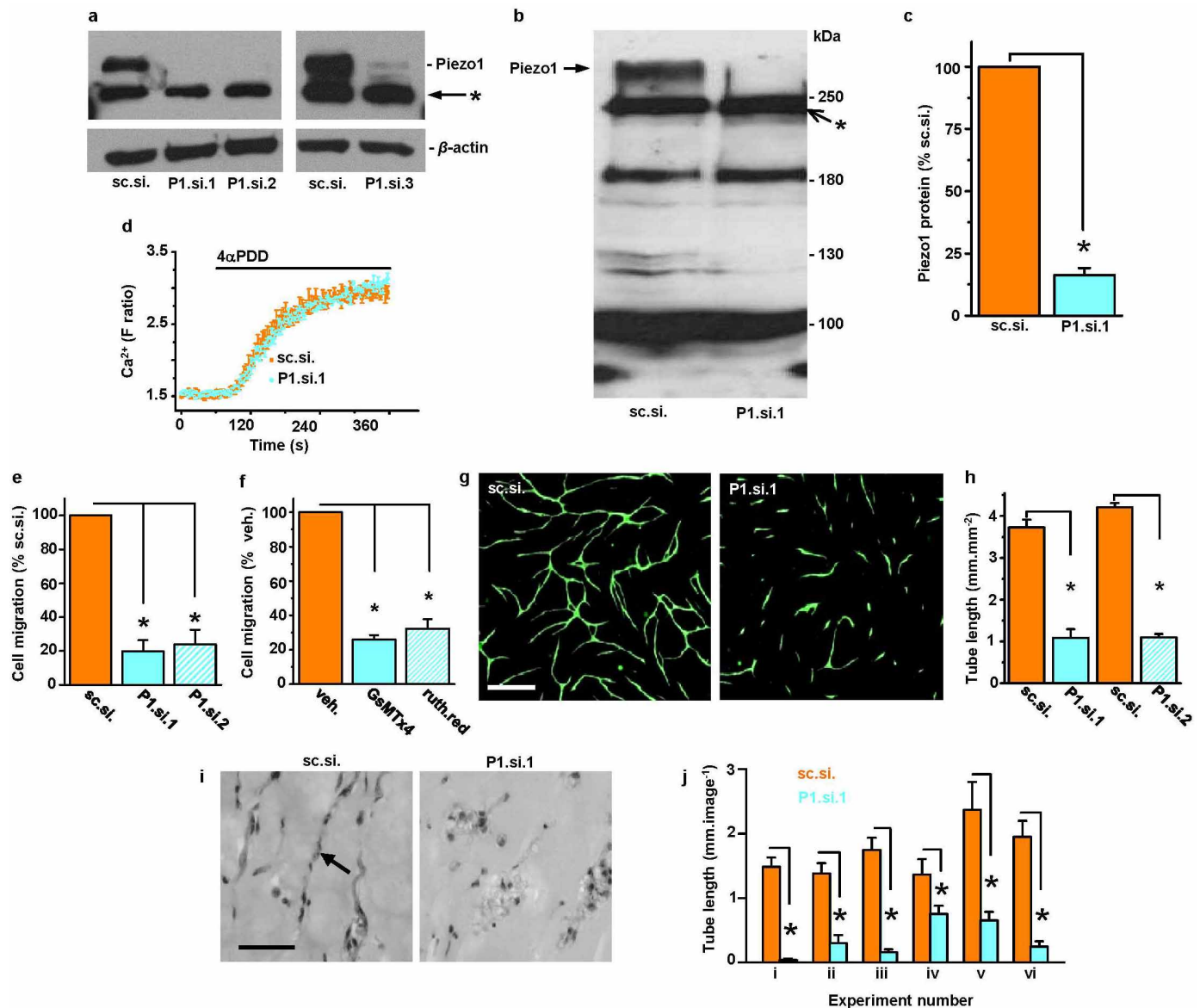


Extended Data Figure 1 | *Piezo1* mRNA in aorta and endothelial cells.

a, End-point PCR products obtained with *Piezo1* primers for human (h.) liver and mouse (m.) lung endothelial cell (EC) and freshly dissected mouse aorta mRNA after reverse transcriptase reaction (RT) to generate cDNA. **b**, As for **a** but for human late outgrowth endothelial progenitor cells (LEPC) and 7 types of human endothelial cell (art., arterial; micr., microvascular; pul., pulmonary;

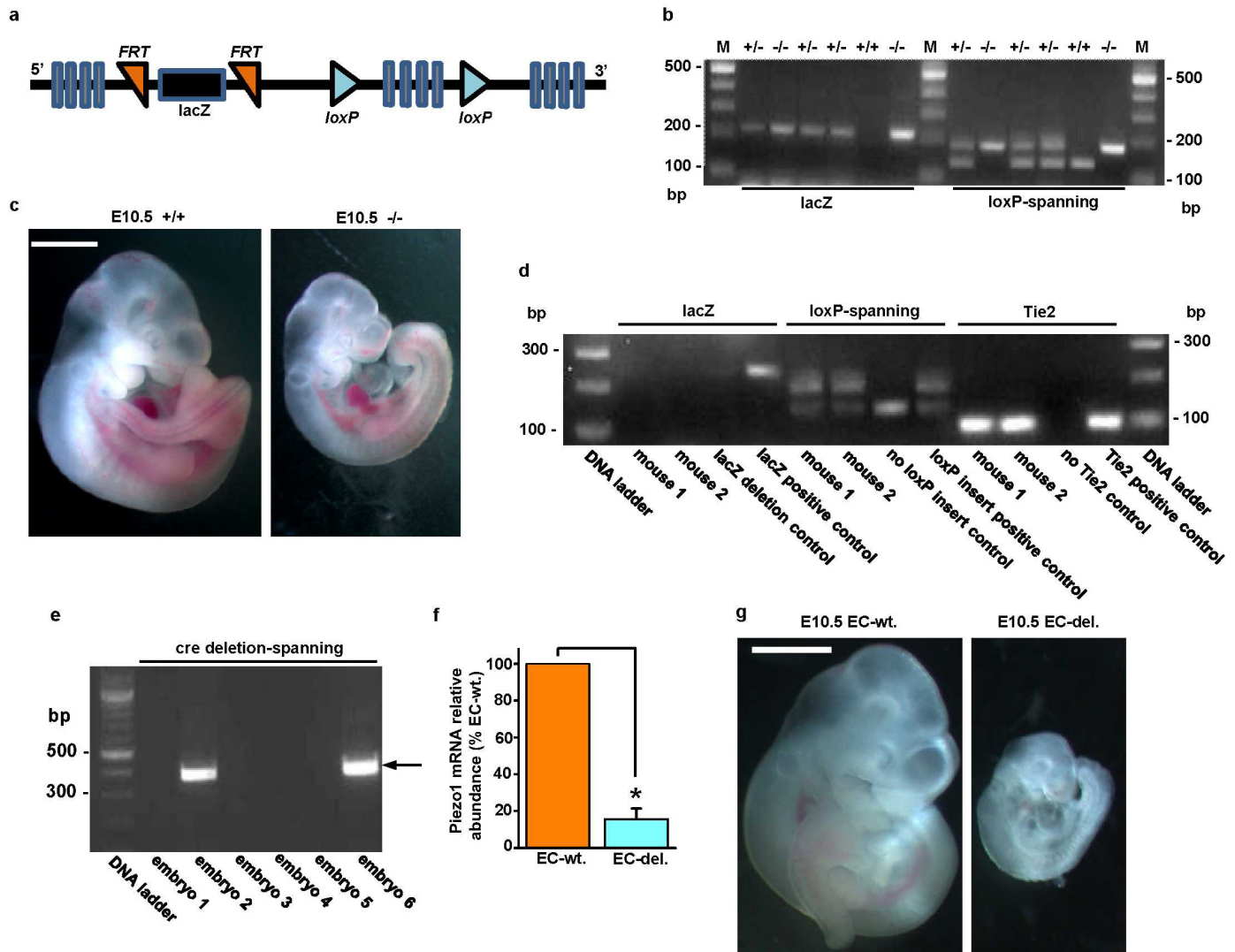
umb., umbilical; car., cardiac; bla., bladder; der., dermal; col., colonic). Results are shown with (+ RT) and without (–RT) reverse transcription.

c, Quantitative real-time PCR data for experiments of the type shown in **a** ($n = 2$ each in duplicate). **d**, Quantitative real-time PCR data for experiments of the type shown in **b** ($n = 1$ each in duplicate).



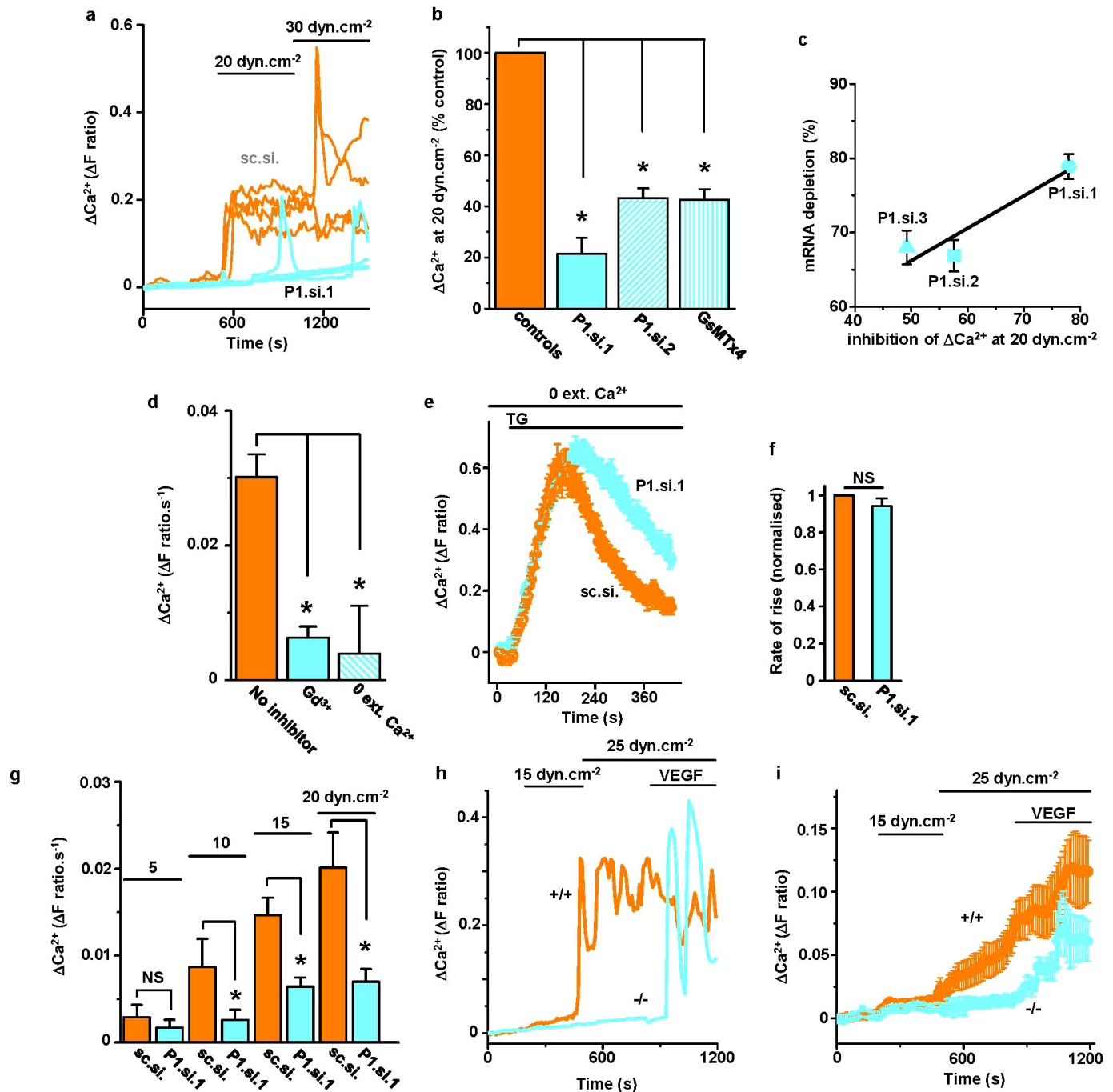
Extended Data Figure 2 | Role of Piezo1 channels in HUVEC migration and tube formation. **a**, Western blot for HUVEC lysate probed with anti-Piezo1 antibody after transfection with the control siRNA (sc.si.), 2 different single Piezo1 siRNAs (P1.si.1 or P1.si.2), or a pooled set of Piezo1 siRNAs (P1.si.3). The upper band in the upper blot represents Piezo1 with a predicted mass of 286 kDa. The band immediately below is unknown protein labelled non-specifically by anti-Piezo1 antibody (*). The lower blot shows β -actin included as a protein-loading control. **b**, Another western blot for HUVEC lysate probed with anti-Piezo1 antibody after transfection with control siRNA (sc.si.) or Piezo1 siRNA (P1.si.1). The arrow points to Piezo1 protein. Apparent depletion of an additional protein by P1.si.1 is evident at about 130 kDa but this effect was not reproducible in other experiments. Other proteins (for example, at about 250(*), 190, and 100 kDa) were non-specifically labelled by the anti-Piezo1 antibody and not affected by Piezo1 siRNA. More specific anti-Piezo1 antibody could not be found. **c**, Normalized quantitative densitometry analysis for Piezo1 band of the type shown in **b** ($n = 3$).

d, Specificity of depletion by Piezo1 siRNA. The effect of the TRPV4 channel activator 10 μ M 4 α -phorbol 12,13-didecanoate (4- α PDD) is shown on intracellular Ca^{2+} in HUVECs in multiple wells of a 96-well plate on a fluorescence plate-reader (representative of $n = 3$). HUVECs were transfected with sc.si. or P1.si.1. The data show that P1.si.1 did not affect TRPV4. **e**, Cell migration after transfection with sc.si. compared with P1.si.1 or sc.si. compared with P1.si.2 ($n = 4$ each). **f**, As for **e** but comparing vehicle controls with 5 μ M GsMTx4 or 30 μ M ruthenium red ($n = 3$ each). **g**, Example images of *in vitro* tube formations in co-culture with fibroblasts. HUVECs were transfected with sc.si. or P1.si.1 and labelled with anti-CD31 antibody (green). Scale bar, 400 μ m. **h**, Analysis of tube length in images of the type shown in **g** and similarly for P1.si.2 ($n = 3$ for all groups). **i**, Example sections from *in vivo* Matrigel plugs in which HUVECs were transfected with sc.si. or P1.si.1. The arrow points to a typical tube structure. Scale bar, 50 μ m. **j**, Mean data from tubes exemplified by **i** for 6 independent experiments (i–vi) (5–17 tissue sections each). Error bars are s.e.m.



Extended Data Figure 3 | Global and endothelial-specific *Piezo1* modification and embryonic growth retardation in mice. **a**, Simplified diagram of the *Piezo1* Knockout First (conditional) construct provided in ES cells by the KOMP Repository (<http://www.komp.org>). *Piezo1* is indicated containing the insertion of *lacZ* sequence flanked by flippase recognition target (FRT) sites and downstream *loxP* sites. Further details of the construct can be obtained at (<http://www.komp.org>). **b**, **c**, Global modification. **b**, Example genotyping results with *lacZ* or *loxP*-spanning PCR primers. M indicates the DNA marker ladder. On the left are results for 6 mice analysed by the *lacZ* PCR primers (expected product: 225 bp). On the right are the results for the same 6 mice analysed by primers targeted to endogenous *Piezo1* sequence either side of the 3' terminal *loxP* site (expected products: 155 bp without the *loxP* site; 189 bp with the *loxP* site). In the gel shown, 3 mice were heterozygous for the construct ($+/-$), 2 homozygous ($-/-$), and 1 wild type ($+/+$). **c**, Images of example sibling E10.5 embryos. The embryo on the left was *Piezo1* $^{+/+}$ and the embryo on the right was *Piezo1* $^{-/-}$. Scale bar, 1 mm. **d–g**, Endothelial-specific modification. **d**, Example genotyping results for two mice (mouse 1 and mouse 2) both with deletion of the *lacZ* insert and transmission of *Tie2*-Cre. Controls for the absence and presence of *lacZ*, the *loxP* insert, and *Tie2*-Cre are included. Successful deletion of the *lacZ* insert was confirmed by lack of

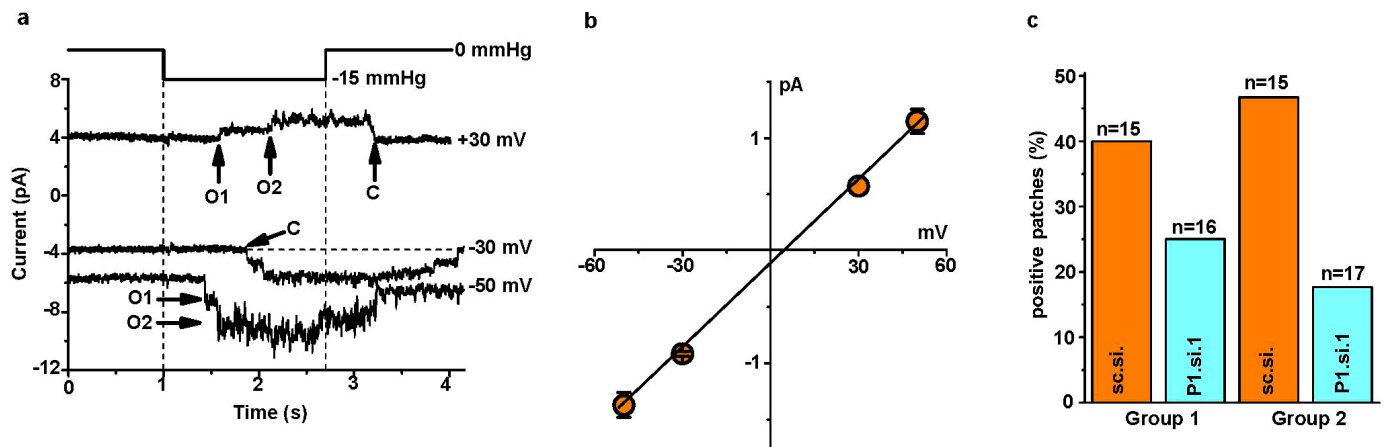
β -galactosidase staining (data not shown). **e**, Example genotyping results for six sibling embryos analysed with PCR primers spanning the deletion predicted to result from Cre recombinase activity at the *loxP* sites. The forward primer was 5' of the 5' FRT site illustrated in **a** and the reverse primer was 3' of the 3' *loxP* site. The PCR product size after deletion was expected to be 379 bp. The product was detected in embryos 2 and 6. The PCR product was not generated in embryos without the deletion because it was too long to be amplified (4,208 bp). Embryos exhibiting the 379 bp product were designated 'EC-del.' to indicate disruptive deletion in *Piezo1* of endothelial cells (ECs). Embryos designated as wild type (wt.) exhibited no 379 bp product and only the 155 bp *loxP* product (as shown for the 'no *loxP* insert control' in **d**). Out of a total of 142 embryos, 57 were EC-del. **f**, RT-PCR products detecting *Piezo1* mRNA in total RNA from sibling embryos (*Piezo1* 3' PCR primers) ($n = 3$, each in duplicate). *Piezo1* mRNA was significantly depleted in embryos displaying the 379 bp product described and shown in **e**. **g**, Images of example sibling E10.5 embryos. The embryo on the left was wild type and the embryo on the right contained the endothelial-specific *Piezo1* deletion (EC-del.). Retarded growth was apparent in EC-del embryos and none of the other embryos. Scale bar, 1 mm. Error bars are s.e.m.



Extended Data Figure 4 | Piezo1-dependence of shear-stress-evoked Ca^{2+} events in human endothelial cells and mouse embryonic endothelial cells.

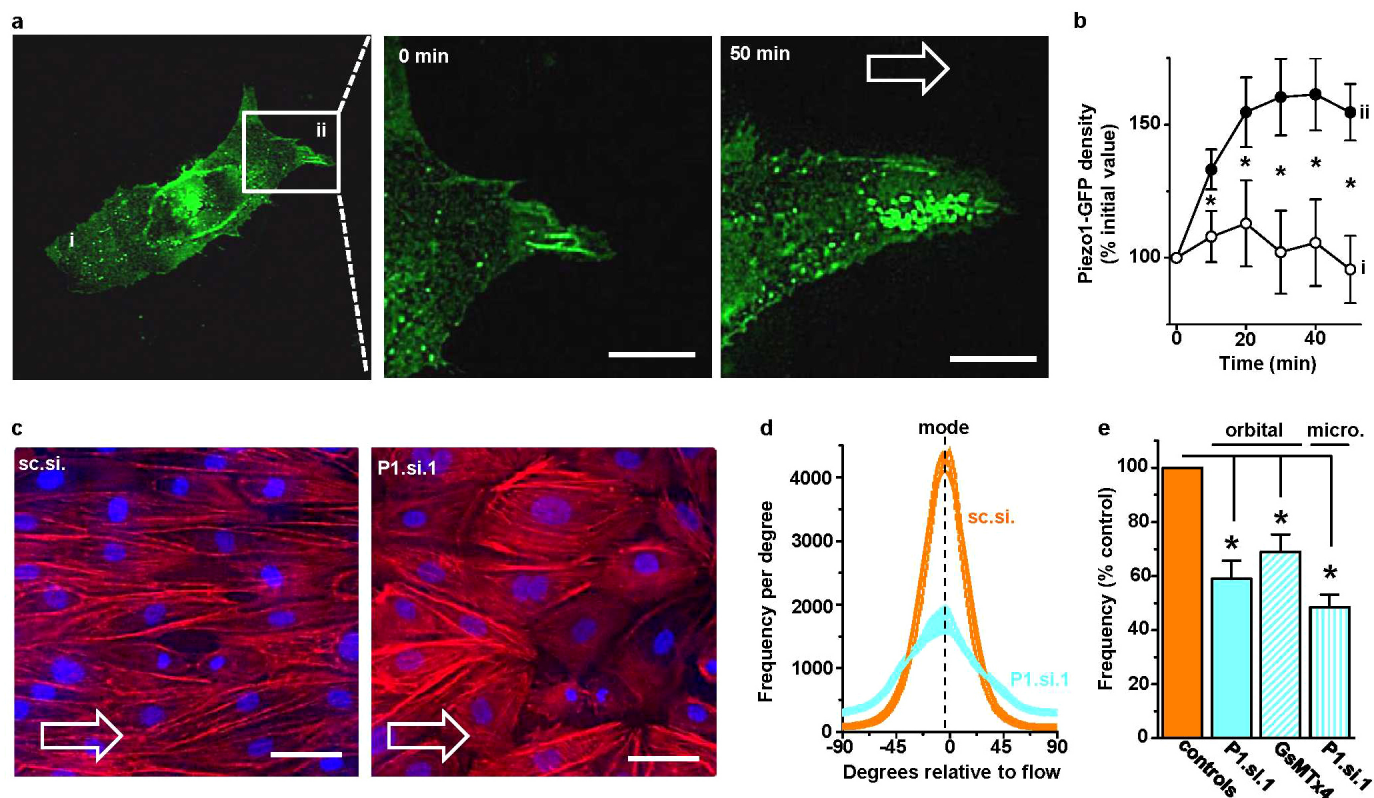
a, Example intracellular Ca^{2+} events evoked by microfluidic shear stress in HUVECs transfected with control siRNA (sc.si.) or Piezo1.si.1 (P1.si.1). Each trace is for 1 cell. In one P1.si.1 cell, transient Ca^{2+} elevation remained. Such residual events may reflect insufficient Piezo1 depletion in some cells or non-Piezo1 mechanisms. **b**, Mean data for experiments of the type in **a** and expanded to paired comparisons of sc.si. and P1.si.1 ($n = 5$ each), sc.si. and P1.si.2 ($n = 4$ each), vehicle and 2.5 μM GsMTx4 ($n = 3$ each). Data were normalized to their respective controls. **c**, Quantification of *Piezo1* mRNA depletion ($n = 4$ each) plotted against the inhibition of the intracellular Ca^{2+} elevations evoked by 20 dyn per cm². Three different Piezo1 siRNAs were compared with their control siRNAs. The Ca^{2+} data are from the experiments described in **b**. Sequence details of the siRNAs are provided in Supplementary Table 3. **d**, Mean Ca^{2+} signals evoked by 20 dyn per cm² in non-transfected HUVECs. Measurements were made in standard bath solution without the addition of an inhibitor (no inhibitor) ($n = 8$), 10 μM gadolinium chloride

(Gd^{3+}) ($n = 3$), or with Ca^{2+} omitted from the bath solution (0 Ca^{2+}) ($n = 3$). **e**, Ca^{2+} release evoked by 2 μM thapsigargin (TG) in the absence of extracellular Ca^{2+} and after transfection with sc.si. or P1.si.1 (20 wells of a 96-well plate each). **f**, Mean data normalized to sc.si. for experiments of the type shown in **e** and analysed for the rate of rise of the Ca^{2+} event evoked by TG ($n = 3$ each). **g**, Similar to **b** but endothelial cells were from patient liver samples, data were not normalized, and only P1.si.1 was used ($n = 3, 4, 10$ and 5 for shear stresses of 5, 10, 15 and 20 dyn per cm²). **h**, **i**, Intracellular Ca^{2+} measurements from mouse embryonic endothelial cells in microfluidic chambers. **h**, Superimposition of example intracellular Ca^{2+} events in 2 single cells on different coverslips from *Piezo1*^{+/+} and *Piezo1*^{-/-} sibling embryos. Shear stress was applied at 15 and 25 dyn per cm² and then 30 ng ml⁻¹ VEGF was introduced while maintaining shear stress at 25 dyn per cm². **i**, Mean \pm s.e.m. data for all VEGF-responsive cells studied as exemplified in **h** ($n = 6$ ^{+/+}, 54 cells; $n = 5$ ^{-/-}, 42 cells). The same data are summarized in simplified form in Fig. 2a. Error bars are s.e.m.



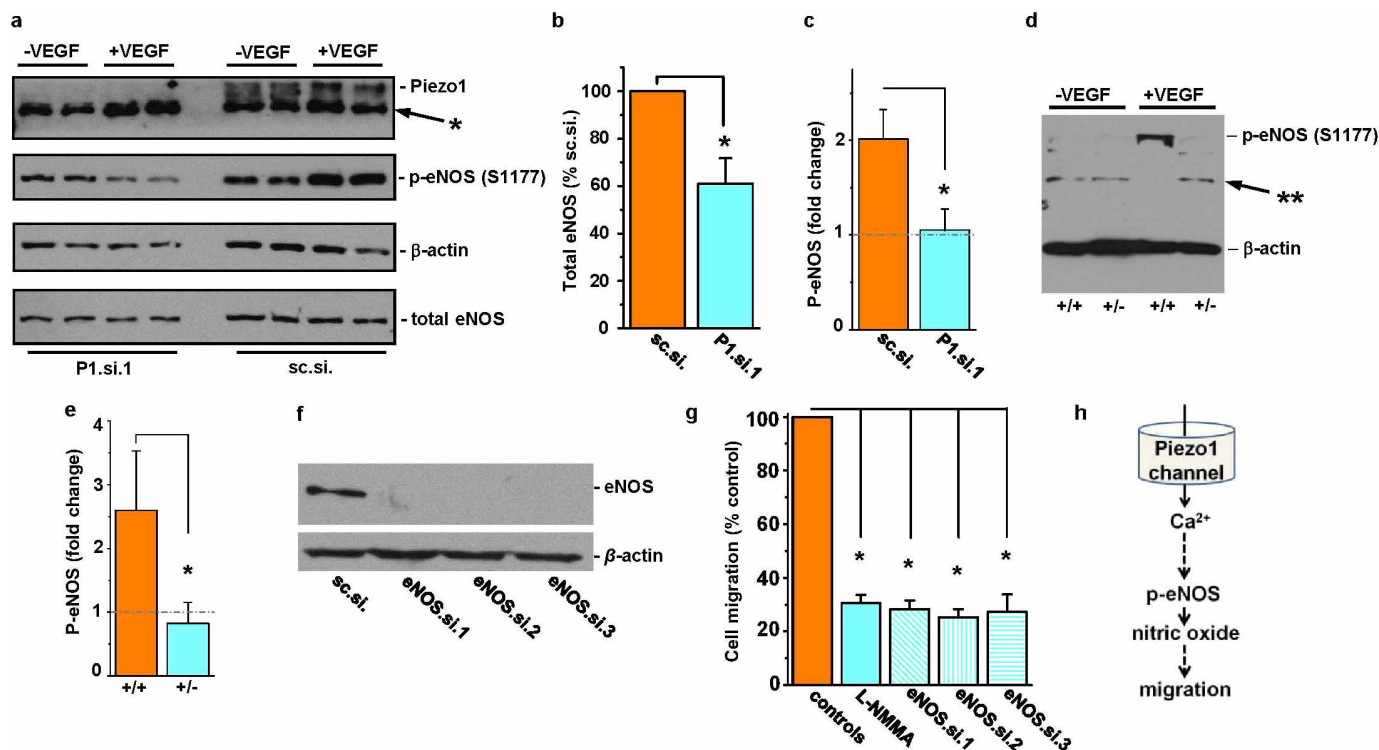
Extended Data Figure 5 | Piezo1-dependence of mechanically activated single channels in HUVECs. **a**, Example single channel currents in a cell-attached patch at three voltages without subtraction of holding current. Application of -15 mm Hg pressure steps to the patch pipette evoked open channel unitary currents that summated to two levels marked as O1 and O2. Closed channel current is indicated by C. **b**, Mean amplitudes of unitary events as exemplified in **a** and fitted with a straight line (3 patches for -50 , -30 and -50 mV; 1 patch for $+30$ mV). **c**, Paired comparisons of the percentage of patches containing channel events exemplified in **a** for cells transfected with

sc.si. or P1.si.1 in two independent experiment groups (n values for each group are in parentheses). In Group 2 cell-attached patch recordings cells were exposed for 10 min to 0.4 mM EGTA to chelate contaminating Ca^{2+} before recording so that sc.si.- and P1.si.1-treated cells rounded up similarly; without this treatment (Group 1), P1.si.1 but not sc.si. cells tended to round up in response to the high- K^+ bath solution used to null the membrane potential of cells in cell-attached patch recordings (the reason for this effect is unknown but it may relate to changes in cytoskeleton and adhesion as discussed in relation to Fig. 4). Error bars are s.e.m.



Extended Data Figure 6 | Shear-stress-evoked redistribution of Piezo1 and the role of Piezo1 in alignment of endothelial cells to the direction of shear stress. The application and direction of shear stress is indicated by open arrows and the cells were HUVECs. **a**, The left-hand image is of Piezo1-GFP in a single cell with a box indicating the region expanded in the middle and right-hand images after 0 and 50 min 15 dyn per cm² in the microfluidic chamber. In the left image i indicates the part of the cell that became trailing after application of shear stress and ii that which became leading. Scale bars, 10 μ m. **b**, Analysis of experiments of the type shown in **a** ($n = 8$ per data point except for $n = 7$ at 50 min) where i and ii indicate the trailing and leading edges

of the cell as shown in **a**. **c**, Example cells after 24 h shear stress caused by the orbital shaker. Rhodamine phalloidin labelled F-actin (red) and DAPI labelled cell nuclei (blue). A paired comparison was made of cells transfected with control siRNA (sc.si.) or Piezo1 siRNA (P1.si.1). Scale bars, 50 μ m. **d**, Example orientation analysis for pairs of images of the type shown in **c**. **e**, As for **d** but normalized mean data for the frequency (number of angles) at the mode in experiments comparing mock with P1.si.1 transfected cells ($n = 5$ each) and 2.5 μ M GsMTx4 with its vehicle control ($n = 4$ each). There is also comparison of cells transfected with sc.si. or P1.si.1 after 15 h of 15 dyn per cm² in the microfluidic chamber ($n = 3$). Error bars are s.e.m.



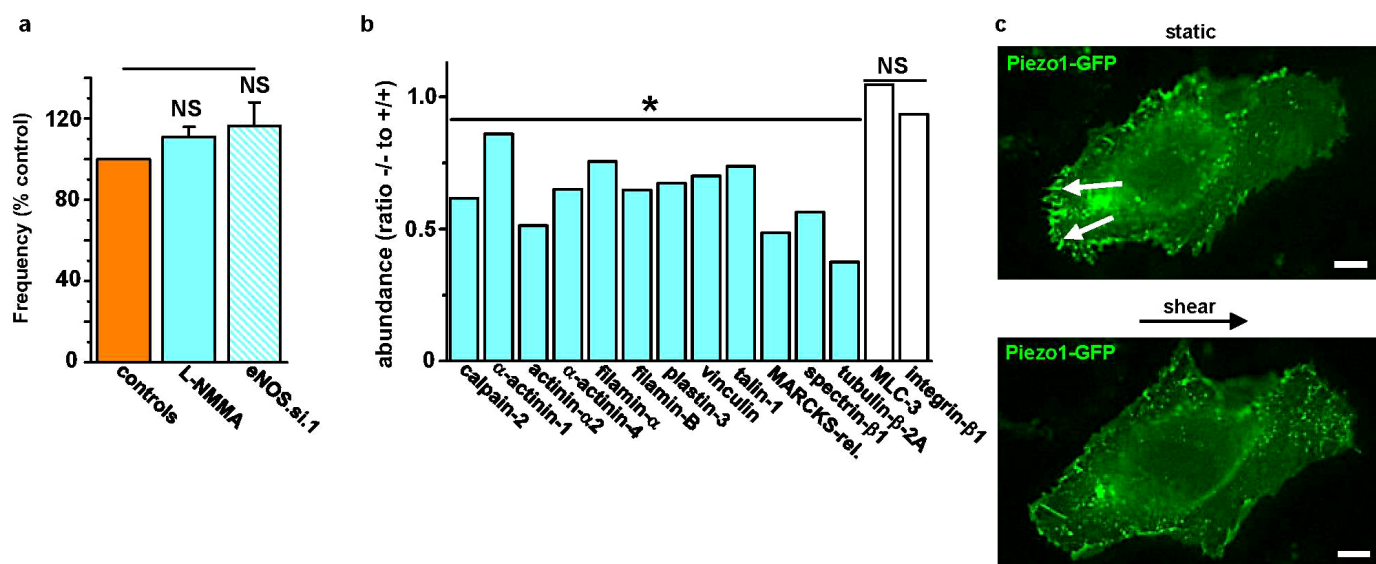
Extended Data Figure 7 | Coupling to endothelial nitric oxide synthase.

a, Western blot for HUVEC lysates probed with anti-Piezo1 antibody after transfection with Piezo1 siRNA P1.si.1 (on the left) or the control siRNA sc.si. (on the right). Prior to collection of cell lysates, HUVECs were treated with 30 ng mL⁻¹ VEGF (+) or no VEGF (-) for 10 min. The lysate was probed with anti-Piezo1 antibody, antibody to phosphorylated S1177 in eNOS, anti-β-actin antibody, and antibody to total eNOS protein. Positions of the expected proteins are indicated by the text on the right. The non-specific band at 250 kDa in the anti-Piezo1 blot is highlighted with *, as in Extended Data Fig. 2a, b.

b, Quantitative data for the downregulation of total eNOS after transfection of HUVECs with P1.si.1 (*n* = 6). **c**, Fold-change in S1177 eNOS phosphorylation (p-eNOS) evoked by VEGF (30 ng mL⁻¹) in HUVECs transfected with control siRNA (sc.si.) or Piezo1 siRNA (P1.si.1) (*n* = 3 each). The grey dashed line highlights 1-fold (that is, no change). **d**, Western blot for VEGF (30 ng mL⁻¹)

evoked S1177 eNOS phosphorylation (arrow) in aorta. Aorta was dissected from *Piezo1*^{+/+} or *Piezo1*^{+/-} litter-mates and allowed to equilibrate at 37°C in culture medium without shear stress for 3 h. Aorta was then exposed to VEGF (30 ng mL⁻¹) (+VEGF) or not (-VEGF) for 10 min, after which lysates were generated. Proteins were probed with antibody to phosphorylation at S1177 in eNOS. The band labelled with ** was not included in the analysis. The blot was also probed with anti-β-actin antibody to test for equal protein loading.

e, Mean data for the type of experiment exemplified in **d** (*n* = 5 for each genotype) and presented as in **c**. **f**, Western blotting for HUVEC lysates after transfection with control siRNA (sc.si.) or eNOS siRNAs. The blot was probed with anti-eNOS (total) antibody. **g**, HUVEC migration to VEGF after incubation with vehicle control, 0.3 mM L-NMMA for 0.5 h, or 48 h after transfection with sc.si. or one of three siRNAs targeted to eNOS (*n* = 3 each; each paired to its own control). **h**, Data interpretation. Error bars are s.e.m.



Extended Data Figure 8 | Endothelial cell alignment to shear stress lacks dependency on nitric oxide but is coupled to calpain. **a**, Frequency of HUVEC alignment induced on the orbital shaker. Data for each test condition were normalized to their own control. Test conditions were 0.3 mM L-NMMA ($n = 3$) and transfection with eNOS siRNA (eNOS.si.1) compared with mock transfection ($n = 4$). **b**, Protein abundances from mass spectrometry analysis for the indicated proteins in 3 *Piezo1*^{-/-} relative to 3 *Piezo1*^{+/+} E10.5

embryos. Calpain-2 and its substrates were less in *Piezo1*^{-/-} embryos. The effects were relatively specific because more than 1,300 of the detected proteins were unchanged by Piezo1 depletion; data for 2 examples (myosin light chain-3 and integrin- β 1) are shown. **c**, Fluorescence images of Piezo1-GFP in a HUVEC before (upper image) and after (lower image) 15 dyn per cm² for 50 min. The small solid arrows point to focal adhesion structures at the trailing edge of the cell. Scale bar, 10 μ m. Representative of $n = 4$. Error bars are s.e.m.

The complete structure of the large subunit of the mammalian mitochondrial ribosome

Basil J. Greber^{1*}, Daniel Boehringer^{1*}, Marc Leibundgut^{1*}, Philipp Bieri¹, Alexander Leitner², Nikolaus Schmitz¹, Ruedi Aebersold^{2,3} & Nenad Ban¹

Mitochondrial ribosomes (mitoribosomes) are extensively modified ribosomes of bacterial descent specialized for the synthesis and insertion of membrane proteins that are critical for energy conversion and ATP production inside mitochondria¹. Mammalian mitoribosomes, which comprise 39S and 28S subunits², have diverged markedly from the bacterial ribosomes from which they are derived, rendering them unique compared to bacterial, eukaryotic cytosolic and fungal mitochondrial ribosomes^{3–5}. We have previously determined at 4.9 Å resolution the architecture of the porcine (*Sus scrofa*) 39S subunit⁶, which is highly homologous to the human mitoribosomal large subunit. Here we present the complete atomic structure of the porcine 39S large mitoribosomal subunit determined in the context of a stalled translating mitoribosome at 3.4 Å resolution by cryo-electron microscopy and chemical crosslinking/mass spectrometry. The structure reveals the locations and the detailed folds of 50 mitoribosomal proteins, shows the highly conserved mitoribosomal peptidyl transferase active site in complex with its substrate transfer RNAs, and defines the path of the nascent chain in mammalian mitoribosomes along their idiosyncratic exit tunnel. Furthermore, we present evidence that a mitochondrial tRNA has become an integral component of the central protuberance of the 39S subunit where it architecturally substitutes for the absence of the 5S ribosomal RNA, a ubiquitous component of all cytoplasmic ribosomes.

Our previous analysis of the porcine 39S mitoribosomal large subunit at 4.9 Å resolution showed the overall fold of the mitoribosomal 16S rRNA as well as the localization of seven mitoribosomal-specific proteins⁶. However, proteins and protein extensions for which no homology models could be generated could not be modelled at this resolution. Additionally, owing to the extensive differences between yeast and mammalian mitoribosomes, the recently reported high-resolution structure of the yeast mitoribosomal large subunit⁵ is of limited use for understanding the mammalian-specific aspects of mitoribosomal structure and function^{4,7}.

Cryo-electron microscopy (cryo-EM) data of porcine 55S mitoribosomes acquired on a movie-mode-enabled direct electron detector combined with movie frame realignment to compensate for beam-induced specimen motion⁸ and maximum-likelihood-based image classification and alignment⁹ yielded a three-dimensional (3D) reconstruction of the 55S mitoribosome (Extended Data Fig. 1a, b) at 3.6 Å resolution (Fourier shell correlation (FSC) = 0.143, 'gold standard'). However, owing to differences in local resolution (Extended Data Fig. 1c), the quality of the density in the 28S subunit part of the cryo-EM map (28S subunit resolution 4.1 Å) was insufficient for reliable model building and refinement. Therefore, we focused the refinement on the 39S subunit, resulting in an improved 3D reconstruction of the 39S subunit at 3.4 Å resolution (Extended Data Fig. 2), suitable for *de novo* model building, structure refinement and validation.

We were able to build and refine the atomic coordinates of 95% of the 16S rRNA nucleotides and 50 of the 52 proteins of the 39S subunit

(Fig. 1a, b, Extended Data Figs 3 and 4 and Supplementary Table 1). The proteins uL1m and bL7m/bL12m (nomenclature according to ref. 10) form the highly flexible L7/L12 and L1 stalks and are therefore not visible in our maps. The proposed total count of 52 proteins includes mL65 (also called MRPS30) and mL66 (MRPS18A), previously assigned to the 28S small subunit but now identified as components of the 39S subunit (Extended Data Fig. 5a–i, Supplementary Discussion and Extended Data Table 1), in addition to 50 of the 51 proteins previously assigned to the 39S subunit^{4,11}. We found no evidence for the presence of MRPL56 in our mitoribosomal sample by mass spectrometry (Supplementary Table 2). All our protein placements at lower resolution⁶ were confirmed by the high-resolution structure, validating our approach that combined cryo-EM with chemical crosslinking and mass spectrometry (CX-MS)¹².

The mammalian mitoribosome contains 11 additional proteins that do not occur in the yeast 54S large mitoribosomal subunit (Extended Data Fig. 4). Several of these proteins localize to functionally important areas, such as the L7/12 stalk base and the central protuberance (CP) (Extended Data Fig. 5j–l), while others are located on the solvent-exposed side of the subunit (Fig. 1c, d and Extended Data Fig. 6).

The peptidyl transferase centre (PTC) of the ribosome is formed exclusively by rRNA¹³ and catalyses peptide bond formation¹⁴. The structural similarity of the rRNA elements in the PTC active site of the mammalian mitoribosome to those in both bacterial and eukaryotic cytosolic ribosomes indicates that the mechanism of the catalytic core of the mammalian mitoribosome is unchanged, consistent with the high level of sequence conservation in this region of the rRNA^{13,15,16} (Fig. 2). In bacterial ribosomes, bL27 is the only protein that approaches the ribosomal active site¹⁵. The density for the very amino terminus of bL27m, although weak in our cryo-EM map, indicates that bL27m also approaches the PTC. We do not observe any density for additional protein elements of mitoribosomal proteins reaching near the catalytic site.

The 55S mitoribosomes used for structure determination were purified partially occupied with A- and P-site tRNAs in the PTC, and the subpopulation obtained by computational sorting (Extended Data Fig. 1) and used for structure determination was enriched with particles containing bound tRNAs. The heterogeneous mixture of tRNA species occupying these mitoribosomes precludes building of any specific tRNA species into the density. However, the universally conserved CCA-3' ends, which carry the covalently bound amino acid substrates or the nascent chain in the A- and P-site, respectively, are identical in all tRNAs, and could therefore be unambiguously identified and built into the cryo-EM density near the PTC (Fig. 2b). Their interactions with the 16S rRNA near the PTC are conserved from the 23S rRNA–tRNA interactions in bacterial ribosomes¹⁵. The conformation of nucleotide U1327 near the PTC is similar to the conformation of the corresponding U2506 nucleotide observed in the activated state of the bacterial PTC in the presence of properly accommodated PTC substrates¹⁵, indicating that our tRNA-containing complexes correspond to a functional state of the mitoribosome (Fig. 2c).

¹Department of Biology, Institute of Molecular Biology and Biophysics, Otto-Stern-Weg 5, ETH Zurich, CH-8093 Zurich, Switzerland. ²Department of Biology, Institute of Molecular Systems Biology, Auguste-Piccard-Hof 1, ETH Zurich, CH-8093 Zurich, Switzerland. ³Faculty of Science, University of Zurich, CH-8057 Zurich, Switzerland.

*These authors contributed equally to this work.

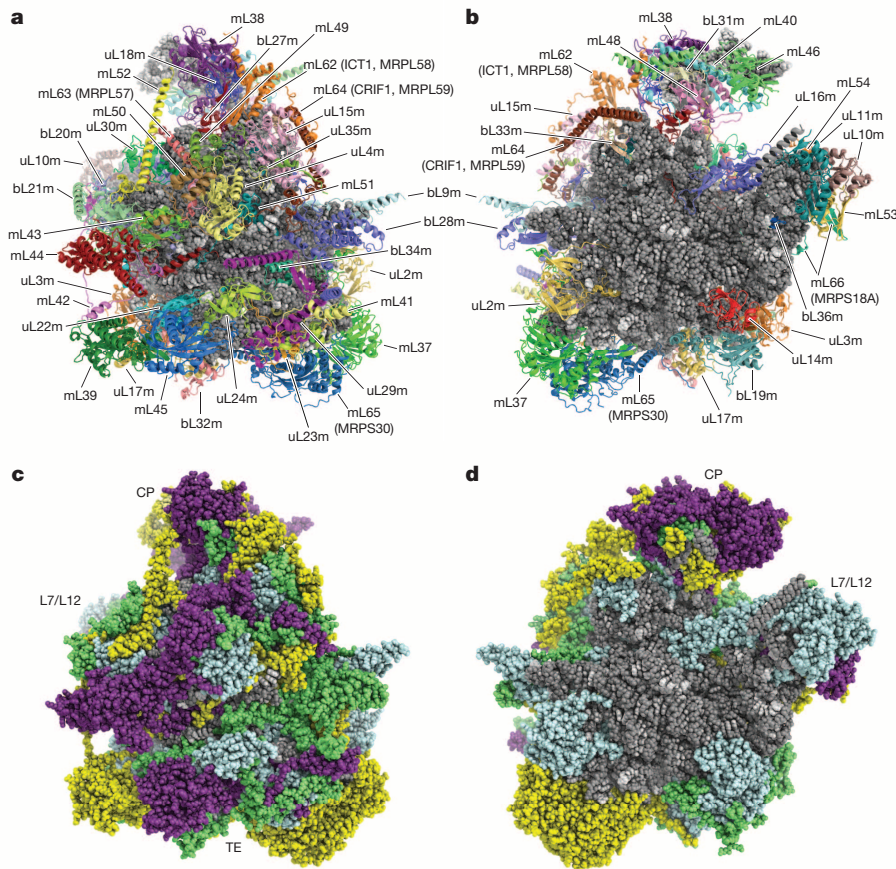


Figure 1 | Overall structure of the 39S subunit. **a, b**, Arrangement of proteins and rRNAs in the 39S subunit (**a**, solvent side; **b**, subunit interface side). **c, d**, 39S subunit (orientation as above) coloured according to the conservation of rRNAs and ribosomal proteins. Grey, 16S rRNA and unassigned protein segments; cyan, proteins conserved in bacteria, mitoribosomal extensions in

lime green; purple, proteins shared between yeast and mammalian mitoribosomes; yellow, CP tRNA and mitoribosomal-specific proteins not present in fungal mitoribosomes. CP, central protuberance; L7/L12, L7/L12 stalk base; TE, tunnel exit.

The polypeptide exit tunnel in the large ribosomal subunit starts at the PTC and leads through the large subunit to the polypeptide exit site (PES), where the nascent protein leaves the ribosome^{13,17}. In bacteria, the polypeptide tunnel is of high medical relevance because it is a binding site for antibiotics¹⁸, whereas the PES region is critical for the processing, folding and targeting of nascent polypeptides¹⁹. Our structure shows that the entry into the mammalian mitoribosomal polypeptide exit tunnel is very similar to the bacterial tunnel, which is the binding site of macrolide antibiotics¹⁸. However, mammalian mitoribosomes are less susceptible

to macrolide antibiotics than bacterial ribosomes²⁰. The structure of the large mitoribosomal subunit from yeast indicated that a constriction of the tunnel near the PTC might be the reason for antibiotic specificity⁵. However, such a constriction is not present in mammalian mitoribosomes (Fig. 3a). Therefore, their resistance to macrolide binding is more likely due to the presence of a G nucleotide at position 1051 (Fig. 3a) of the 16S rRNA (2058 in *Escherichia coli* 23S rRNA), as is the case for eukaryotic cytosolic and archaeal ribosomes^{16,21}. Because of the high homology between human and porcine mitoribosomes and the conservation of G1051

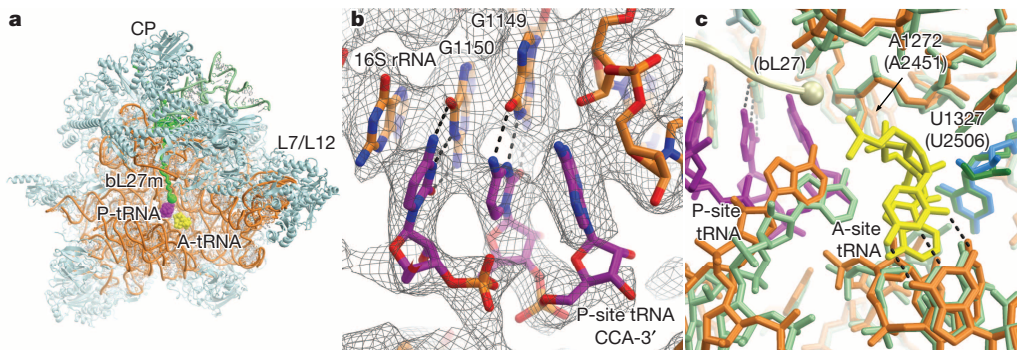


Figure 2 | The peptidyl transferase centre of the 39S subunit. **a**, View into the PTC of the 39S subunit. 16S rRNA in orange; mitoribosomal proteins in cyan; A- and P-site tRNA CCA-3' ends in yellow and purple, respectively; bL27m in bright green. **b**, View of the CCA-3' end of the P-site tRNA bound

near the PTC. Dashed lines indicate base pairing. **c**, Comparison of the bacterial¹⁵ and mitoribosomal PTCs (bacterial structural elements labelled in brackets). Colours as in **a**, with bacterial rRNA in green, bL27 in pale yellow, and base U1327 in dark blue (bacterial U2506 in dark green).

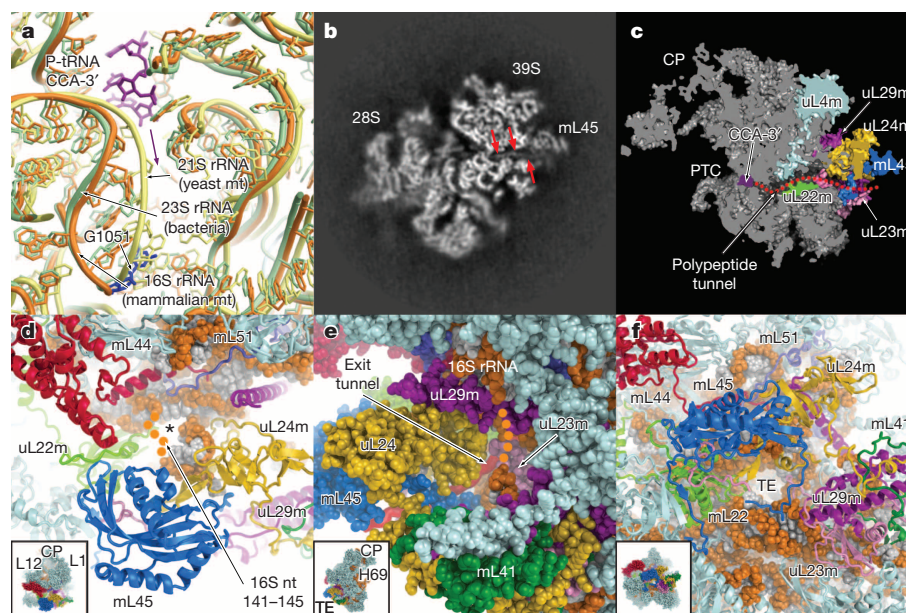


Figure 3 | The mitoribosomal polypeptide exit tunnel. **a**, Comparison of the tunnel entry in mammalian (orange; G1051 dark blue) and yeast (yellow) mitoribosomes and bacterial ribosomes (light green)^{5,15}. P-site tRNA and tunnel entry (arrow) in purple. **b**, Slice through the 39S subunit cryo-EM map (filtered to 5.5 Å resolution). A threadlike density inside the tunnel (red arrows) probably corresponds to the nascent polypeptide. **c**, Schematic representation of the exit tunnel path (same orientation as in **b**). **d**, 16S rRNA nucleotides

141–145 (indicated by orange dotted line because of low local resolution) block the proposed yeast polypeptide tunnel exit site (asterisk)⁵. **e**, View of the polypeptide accessible site (PAS)³, which is reduced to a narrow opening by surrounding proteins and 16S rRNA nucleotides 19–24 (not built, schematically indicated by a dotted line). Exit tunnel path is indicated by a red tube. **f**, View of the canonical polypeptide tunnel exit site (TE) and surrounding proteins.

in the human 16S rRNA²², this mechanism probably also applies to human mitoribosomes.

In mitoribosomes, the strongly remodelled PES region is an important binding site for the membrane insertion machinery, including the Oxa1L insertase^{1,23–25}. In our reconstruction of the 39S subunit, we observe a continuous threadlike density inside the polypeptide tunnel extending from the P-site tRNA in the PTC towards the canonical PES, suggesting that we are observing stalled ribosomes that carry nascent polypeptide chains (Fig. 3b, c). Two alternative exit sites of the mitoribosomal tunnel have been suggested previously: (1) the tunnel of the mammalian mitoribosome bears a lateral opening, termed the polypeptide accessible site (PAS)^{3,25}; and (2) the tunnel of the yeast 54S mitoribosomal large subunit has been suggested to deviate from the otherwise universally conserved path to exit from the ribosome at a location close to proteins mL44 and mL50 (ref. 5). Our data indicate that in mammalian mitoribosomes, the putative exit site observed in yeast mitoribosomes is blocked by the 16S rRNA (Fig. 3d) and that the second proposed lateral exit site, the PAS³, is partially blocked by mitoribosomal-specific protein elements (Fig. 3e). Therefore, based on our structural data on the 39S mitoribosomal subunit and the direct observation of the nascent chain features in our cryo-EM reconstruction, we consider it most likely that mammalian mitoribosomes use the canonical nascent polypeptide exit site (Fig. 3f).

On the basis of rRNA sequence analysis and lower-resolution cryo-EM reconstructions, it has been suggested that the E-site of the mammalian 39S subunit is weak or completely absent^{3,25}. Indeed, in contrast to the yeast 54S mitoribosomal subunit⁵, our structure does not show density for E-site tRNA and confirms that a number of rRNA elements interacting with the E-site tRNA in bacterial ribosomes, such as H68, are absent (Extended Data Fig. 7a, b). However, our structure reveals that the binding pocket for the CCA-3' end of the E-site tRNA observed in bacteria^{15,26} is well conserved in the 39S subunit and in a similar conformation compared to the occupied bacterial E-site (Extended Data Fig. 7c). This makes it likely that the tRNAs in mammalian mitoribosomes move through the ribosome by binding to the E-site on the 39S subunit according to the same mechanistic principles as established for cytoplasmic ribosomes^{27,28}.

Even though the basic mechanics of tRNA translocation are probably conserved, the mammalian mitoribosome may support accurate tRNA binding or translocation with mitoribosomal-specific protein elements, such as the P-site finger, which reaches near the P-site tRNA³. However, the details of these interactions are still poorly understood. We see EM density for two protein elements that approach the area where the tRNAs are located, although at lower local resolution than the surrounding area (Extended Data Fig. 7d–f). At the CP, mL48 and mL40 may form a portion of the P-site finger (Extended Data Fig. 7d), which is inserted between the A- and P-site tRNAs and contacts both tRNAs (Extended Data Fig. 7e, f). Additionally, a long α -helix of mL64 (CRIF1, MRPL59) that emanates from the L1 stalk region of the subunit projects towards the P-site finger and may also contribute to protein–RNA interactions in this region (Extended Data Fig. 7d).

We have previously demonstrated the presence of a second ribosomal RNA molecule in the CP of the 39S subunit in a position similar to a domain of 5S rRNA in bacterial ribosomes⁶. The CP is better ordered in the 55S mitoribosome due to intersubunit contacts to the head domain of the 28S subunit that probably involve mL46 and mL48 (Extended Data Fig. 5j, k), revealing that the RNA density has the characteristic L-shape of a tRNA (Fig. 4a). Without any previous assumptions regarding the identity of this RNA, we built the sequence of purines and pyrimidines into the better-resolved part of the cryo-EM density in the area tentatively assigned as the anticodon stem of the tRNA (Extended Data Fig. 8a–e). Alignment of this pattern to the porcine mitochondrial genome as well as the sequences of all cytosolic tRNAs and rRNAs identified mitochondrial tRNA^{Phe} as the most likely candidate^{29,30}. We therefore conclude that a tRNA, possibly mitochondrial tRNA^{Phe}, is present in the CP of the 39S subunit. We refer to it as CP tRNA to distinguish it from tRNAs used for decoding in the intersubunit space.

The CP tRNA acts as an organizing scaffold for the 39S CP by forming extensive contacts to the surrounding proteins uL18m, mL38, mL40 and mL48 (Fig. 4b). Prokaryotic uL18 binds to domain β of the 5S rRNA¹³, while mitochondrial uL18m binds the CP tRNA near the anticodon stem loop in mammalian mitoribosomes (Fig. 4b). uL18 is absent from the yeast mitoribosome, which lacks both 5S rRNA and CP tRNA, but

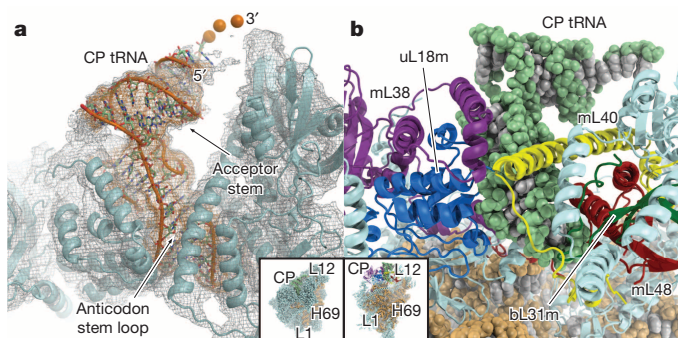


Figure 4 | The CP tRNA in the 39S subunit. **a**, The density for the CP tRNA (orange, map filtered to 6.5 Å resolution) exhibits the L-shape characteristic of tRNAs (mitoribosomal proteins are shown in cyan; CP tRNA CCA-3' end lacking density shown as spheres). **b**, The CP tRNA (light green) forms extensive contacts to uL18m (blue), bL31m (dark green), mL38 (violet), mL40 (yellow) and mL48 (dark red).

contains an rRNA expansion segment in this area⁵. Other elements of mitoribosomal CP architecture, such as mL38, mL40 and mL46, are conserved between yeast and mammals in spite of these differences in RNA content and structure (Extended Data Fig. 8g–i). The recruitment of an entirely novel RNA molecule highlights the extreme evolutionary plasticity of the mammalian mitoribosome. The high-resolution structure presented here enables future structure-based experiments and paves the way for a detailed molecular understanding of mitoribosomal function and evolution (Supplementary Discussion and Extended Data Fig. 9).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 July; accepted 26 September 2014.

Published online 1 October 2014.

- Ott, M. & Herrmann, J. M. Co-translational membrane insertion of mitochondrially encoded proteins. *Biochim. Biophys. Acta* **1803**, 767–775 (2010).
- O'Brien, T. W. The general occurrence of 55 S ribosomes in mammalian liver mitochondria. *J. Biol. Chem.* **246**, 3409–3417 (1971).
- Sharma, M. R. et al. Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. *Cell* **115**, 97–108 (2003).
- Desmond, E., Brochier-Armanet, C., Forterre, P. & Gribaldo, S. On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. *Res. Microbiol.* **162**, 53–70 (2011).
- Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
- Greber, B. J. et al. Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* **505**, 515–519 (2014).
- Smits, P., Smeitink, J. A. M., van den Heuvel, L. P., Huynen, M. A. & Ettema, T. J. G. Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Res.* **35**, 4686–4703 (2007).
- Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
- Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Ban, N. et al. A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* **24**, 165–169 (2014).
- Koc, E. C. et al. Identification and characterization of CHCHD1, AURKAIP1, and CRIF1 as new members of the mammalian mitochondrial ribosome. *Front. Physiol.* **4**, 183 (2013).
- Walzthoeni, T., Leitner, A., Stengel, F. & Aebersold, R. Mass spectrometry supported determination of protein complex structure. *Curr. Opin. Struct. Biol.* **23**, 252–260 (2013).

- Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
- Beringer, M. & Rodnina, M. V. The ribosomal peptidyl transferase. *Mol. Cell* **26**, 311–321 (2007).
- Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C. & Ramakrishnan, V. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nature Struct. Mol. Biol.* **16**, 528–533 (2009).
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S. & Ban, N. Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* **334**, 941–948 (2011).
- Voss, N. R., Gerstein, M., Steitz, T. A. & Moore, P. B. The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.* **360**, 893–906 (2006).
- Schlünzen, F. et al. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**, 814–821 (2001).
- Boehringer, D., Greber, B. & Ban, N. in *Ribosome Structure, Function, and Dynamics* (eds Rodnina, M., Wintermeyer, W. & Green, R.) 405–418 (Springer, 2011).
- Zhang, L. et al. Antibiotic susceptibility of mammalian mitochondrial translation. *FEBS Lett.* **579**, 6423–6427 (2005).
- Tu, D., Blaha, G., Moore, P. B. & Steitz, T. A. Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. *Cell* **121**, 257–270 (2005).
- Cannone, J. J. et al. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinform.* **3**, 2 (2002).
- Stiburek, L. et al. Knockdown of human Oxa11 impairs the biogenesis of F₁F₀-ATP synthase and NADH:ubiquinone oxidoreductase. *J. Mol. Biol.* **374**, 506–516 (2007).
- Saller, M. J., Wu, Z. C., de Keyser, J. & Driessen, A. J. M. The YidC/Oxa1/Alb3 protein family: common principles and distinct features. *Biol. Chem.* **393**, 1279–1290 (2012).
- Agrawal, R. K. & Sharma, M. R. Structural aspects of mitochondrial translational apparatus. *Curr. Opin. Struct. Biol.* **22**, 797–803 (2012).
- Selmer, M. et al. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
- Moazed, D. & Noller, H. F. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**, 142–148 (1989).
- Agirrezabala, X. et al. Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol. Cell* **32**, 190–197 (2008).
- Jühling, F. et al. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* **37**, D159–D162 (2009).
- Helm, M. et al. Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA* **6**, 1356–1379 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements Cryo-EM data were collected at the electron microscopy facility of ETH Zurich (ScopeM). We thank P. Tittmann (ScopeM) for support. We thank J. Yu and C. Ciaffalo for discussions. This work was supported by the Swiss National Science Foundation (SNSF), the National Center of Excellence in Research (NCCR) Structural Biology program of the Swiss National Science Foundation (SNSF), European Research Council (ERC) grant 250071 under the European Community's Seventh Framework Programme (to N.B.), the Commission of the European Communities through the PROSPECTS consortium (EU FP7 projects 201648, 233226) (R.A.), and the European Research Council (ERC-2008-AdG 233226) (R.A.).

Author Contributions P.B. and B.J.G. performed preparation of the mitoribosomes. B.J.G., P.B. and D.B. prepared cryo-EM samples. D.B. and B.J.G. acquired the cryo-EM data. B.J.G. and P.B. calculated the cryo-EM reconstructions with support from D.B. M.L. and B.J.G. built the atomic model. M.L., N.S. and D.B. performed coordinate refinement of the atomic model. M.L., B.J.G., P.B., D.B. and N.B. interpreted the structure. A.L. performed CX-MS experiments in the laboratory of R.A. All authors contributed to the final version of the paper.

Author Information The 3.4 Å cryo-EM map of the 39S mitoribosomal subunit has been deposited in the Electron Microscopy Databank with accession code EMD-2787. The coordinates of the atomic structure of the 39S mitoribosomal subunit have been deposited in the Protein Data Bank under accession codes 4v1a and 4v19. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.B. (ban@mol.biol.ethz.ch).

METHODS

Sample preparation and data acquisition. Porcine mitochondria and mitochondrial ribosomes were prepared as described⁶. Cryo-EM grids were prepared by applying a sample solution containing 45 nM 55S mitoribosomes in sample buffer (20 mM HEPES-KOH pH 7.4, 40 mM MgCl₂, 50 mM KCl, 1 mM DTT) to Quantifoil R2/2 holey carbon grids (Quantifoil Micro Tools) coated with a thin carbon film (prepared using a Balzers BAE 120 thin-film coating system). The grids were manually blotted with damp filter paper from the front and manually plunged into a 2:3 mixture of ethane and propane³¹ at liquid nitrogen temperature. Cryo-EM data were collected using a Titan Krios cryo-transmission electron microscope (FEI Company) operated at 300 keV equipped with a Falcon II direct electron detector (FEI Company) at a magnification of 100,000 \times (accuracy of magnification calibration approximately 0.5%) and defocus values of -0.8 to -3.0 μ m. The EPU software (FEI Company) was used for automated data acquisition. Images were acquired by collecting seven movie frames (combined dose 20 electrons per \AA^2) per exposure (combined exposure time 0.77 s) after discarding the first frame (55 ms). The movie frames were subsequently aligned to correct for beam-induced specimen motion using DOSEFGPU DRIFTCORR⁸.

Structure calculation. Data was acquired in two sessions, yielding two data sets that were initially processed independently. Overview images of the holes from which data had been collected were acquired by the EPU software and inspected to discard areas of poor ice quality or cracks in the carbon film. From the drift-corrected micrographs, CTF parameters were estimated using CTFFIND³² from within RELION⁹. The quality of the power spectra produced in this step was evaluated and poor quality micrographs were rejected. Roughly 231,000 (data set 1) and 328,000 (data set 2) particles were selected using Batchboxer (EMAN 1.9)³³ as described⁶. Further image processing was performed in RELION 1.2 (ref. 9). After an initial round of 2D-classification with 120 classes at a pixel size of 5.6 \AA on the object scale (80 pixel frames) to remove most of the non-ribosomal particles and 39S subunits, a 3D classification with 10 classes at the same pixel size was performed, using a low-pass filtered reconstruction of the 55S mitoribosome as an initial reference⁶, to sort the particle images into homogeneous subpopulations and again remove particles that were not 55S mitoribosomes (Extended Data Fig. 1a). After pooling of the 55S mitoribosomal particles from the two data sets, a second round of 3D classification with 10 classes was performed. Roughly 142,000 particles were then selected and refined to high-resolution using 'gold standard' refinement in RELION. A 3.6 \AA resolution reconstruction of the 55S mitoribosome was obtained (Extended Data Fig. 1b, c), which however was not of sufficient resolution for *de novo* model building of the 28S subunit part due to low local resolution (4.1 \AA on average for the 28S subunit). Therefore, refinement was limited to the 39S subunit by applying a mask around the subunit as described previously^{5,35}, resulting in a 3.4 \AA resolution reconstruction of the 39S subunit (Extended Data Fig. 2) that could be used for manual atomic model building, refinement and validation.

Chemical crosslinking using *S. scrofa* 55S mitoribosomes. Crosslinking using the amine-reactive disuccinimidyl suberate (DSS) was performed as described previously⁶ using 75 μ g of the 55S complex at a concentration of 0.5 mg ml⁻¹. Combined crosslinking of acidic residues and zero-length crosslinking using pimelic acid dihydrazide (PDH) and 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) was performed following the procedure described recently³⁶. The ribosome sample (75 μ g) was prepared at a concentration of 0.5 mg ml⁻¹ in crosslinking buffer (20 mM HEPES-KOH pH 7.6, 100 mM KCl, 20 mM MgCl₂, 1 mM DTT, 10% (w/v) sucrose). PDH was added as an equimolar mixture of isotopically light (d_0) and heavy (d_{10}) form (both obtained from Sigma-Aldrich) to a final concentration of 9 mg ml⁻¹. The crosslinking reaction was initiated by addition of DMTMM to a final concentration of 12 mg ml⁻¹, and the sample was incubated at 37 $^{\circ}$ C for 45 min. The reaction was stopped by gel filtration using a Zeba spin column (ThermoFisher/Pierce) according to the manufacturer's instructions.

Further sample treatment (reduction, alkylation, enzymatic digestion using endoproteinase Lys-C and trypsin, size exclusion chromatography fractionation) and liquid chromatography-tandem mass spectrometry (LC-MS) analysis (Easy nLC-1000/Orbitrap Elite, ThermoFisher) followed published procedures^{6,37,38}.

MS data were searched using xQuest/xProphet³⁹ as described previously^{6,36} against a database compiled from *S. scrofa* mitoribosomal protein sequences identified by protein-protein BLAST⁴⁰ or BLAST searches of human mitoribosomal protein sequences against a translated nucleotide database (NCBI nucleotide collection, <http://blast.ncbi.nlm.nih.gov/>; limited to *S. scrofa* sequences), as described⁶. A false discovery rate of 5% was set using xProphet, independently for different crosslinking chemistries and for intra- and inter-protein crosslinks.

Analysis of non-crosslinked *S. scrofa* 55S ribosomes using LC-MS/MS. For the assignment of mitoribosomal protein variants (Supplementary Table 2) and label-free relative quantification of 55S subunits (Extended Data Fig. 5i), a 10 μ g aliquot of untreated 55S mitoribosomes was reduced, alkylated and digested with Lys-C and trypsin. 1 μ g of the digest was analysed in duplicate by LC-MS/MS using the

same experimental set-up as for the crosslinked fractions, with the following exceptions: a 120-min gradient from 5% to 30% B, where solvent A = water/acetonitrile/formic acid (95:5:0.1 (v/v/v)) and B = acetonitrile/water/formic acid (95:5:0.1 (v/v/v)), was used, and the 15 most abundant precursors with charge states of 2+ or above were selected for fragmentation.

MS data were converted into the Mascot generic format using MzXML2Search (part of the Trans-Proteomic pipeline⁴¹) and searched against the custom mitoribosomal protein database (see above) using Mascot 2.4.01 (ref. 42). Protein grouping was performed in Mascot as described in ref. 43. Data were extracted manually from the Mascot report, and a minimum subset of proteins was assembled using only database entries for which unique peptide sequences were assigned.

Label-free quantification was performed using Skyline 2.5.0.5675 (ref. 44) based on the database search results exported in Mascot .DAT format. Quantified peak areas provided by Skyline were used to aggregate cumulative peak areas of the top 3 or top 5 identified peptides^{45,46}.

Atomic model building. The excellent experimental cryo-EM density allowed us to manually build an almost complete atomic model of the 39S subunit (Supplementary Table 1) using O^{47,48} (proteins) and COOT⁴⁹ (rRNA). The atomic model was mostly built based on our previous lower resolution structure⁶ or by complete *de novo* tracing of novel protein folds. mL40, mL46 and mL50 were rebuilt based on templates from the yeast 54S subunit structure⁵. We used the sequences of the isoforms identified by LC-MS/MS analysis (Supplementary Table 2) whenever a conclusive assignment could be reached. Three unassigned β -strands were identified at a similar location to bL31m of the yeast 54S structure. A genome-wide BackPhyre search⁵⁰ using the corresponding yeast structure in the four implemented metazoan genomes (human, mouse, *Drosophila melanogaster* and *Caenorhabditis elegans*) identified the mammalian mitoribosomal protein MRPL55 as the closest homologue of bL31m, although with relatively low sequence conservation (<22%). Consequently, we renamed the MRPL55 to bL31m (see below) and corroborated the identity of the protein in the EM density. Remaining unoccupied EM density peaks could be interpreted as coordinated magnesium or zinc ions (zinc finger proteins bL32m, bL36m and mL66 (MRPS18A), the latter forming a mixed zinc finger together with uL10m, which contributes one zinc-coordinating residue). In spite of the high quality of our cryo-EM density, we cannot fully exclude register shifts in poorly ordered or peripheral regions of the structure.

Structure refinement and validation. For the use of atomic coordinate refinement, the final 39S subunit cryo-EM maps from the automask output of RELION were clipped to a unit cell that generously accommodated the entire large subunit (resulting in arbitrary unit cell dimensions of 302.4 \AA \times 302.4 \AA \times 302.4 \AA with $\alpha = \beta = \gamma = 90^{\circ}$). We then refined and validated the structure against reciprocal space data calculated from these cryo-EM maps (see below). The final refinement statistics are given in Extended Data Fig. 3a, and the distribution of B-factors in the refined molecular model is shown in Extended Data Fig. 3b.

For the conversion of the cryo-EM maps to reciprocal space structure factors we used a customized CNS⁵¹ input script, which blurs the phase probabilities of the structure factors in a resolution-dependent manner based on the FSC values to account for the decreasing reliability of the map-derived amplitudes and phases with increasing resolution. For each of the resulting 1,472,132 reflections, a figure of merit (FOM, indicating the probability of the phase angle ϕ to be correct) was calculated at the corresponding resolution⁵² according to equations (1) for the final map calculated from the full data set and equation (2) for the half set maps calculated using data half sets⁵³

$$\text{FOM} = (2 \text{ FSC} / (1 + \text{FSC}))^{1/2} \quad (1)$$

$$\text{FOM} = \text{FSC}^{1/2} \quad (2)$$

For the reciprocal space refinement in PHENIX⁵⁴, experimental phase restraints were applied throughout by refining against the MLHL (phased maximum likelihood) target function after conversion of the FOMs and phase angles to the corresponding Hendrickson-Lattman (HL) coefficients, which represent an alternative way of expressing the phase probabilities. After one round of rigid body refinement (one rigid group for each protein and tRNA, five groups for the 16S rRNA), the entire model was refined to convergence by applying six macrocycles of individual coordinate and B-factor refinement using protein Ramachandran and secondary structure or RNA base-pair restraints to maintain good geometry in less well resolved regions of the map. To account for the specific electron scattering properties, we used the atomic electron scattering table provided by PHENIX.

Overfitting of the resulting model was monitored and minimized by a procedure according to the method described previously^{5,35}. In short, we manually tested different geometry weights during the refinement of the model against the full data set. Models with reasonable geometry values (90% of representative entries of similar

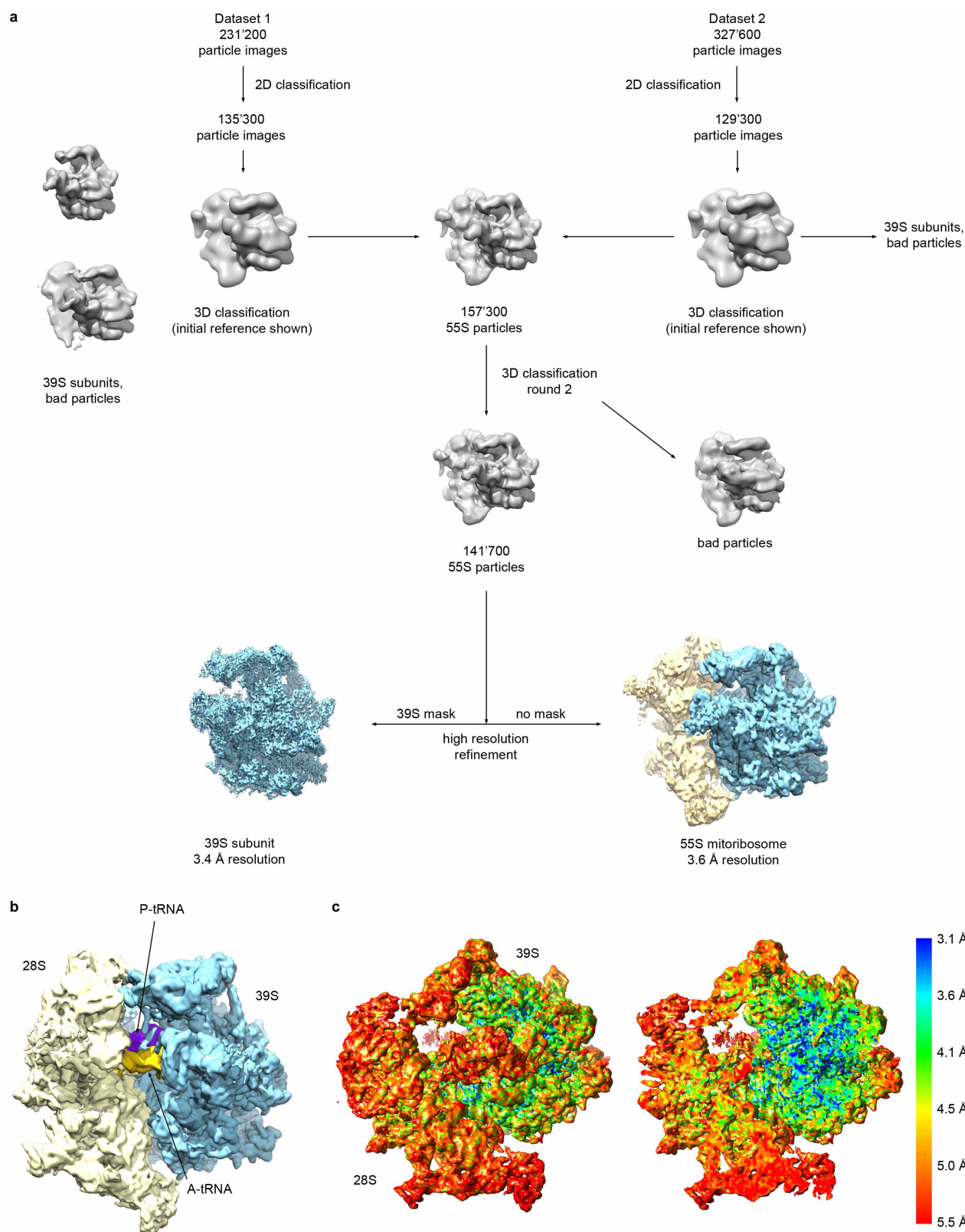
resolution: root mean square deviation (r.m.s.d.) bonds <0.014 Å, r.m.s.d. angles <2.0 degrees, R-factors $<30\%$ were further validated⁵⁵. The coordinates of the models were displaced randomly by 0.1 Å to remove potential model bias and subsequently re-refined to convergence against reciprocal space data generated from one of the half maps used in the final reconstruction. Then we calculated the FSC of the resulting model against each of the two half maps (Extended Data Fig. 3c–e) and compared the decay of the two FSC curves. A substantial gap between the two curves is indicative of overfitting^{5,35}.

In our refinement we determined a weight of 1.0 for the reciprocal space data as the optimal balance between the overall fit to the experimental map and model geometry for our model (Extended Data Fig. 3c). Increasing the weight of reciprocal data during refinement leads to better agreement of the model to the half-map and lower R-factors but also leads to overfitting, indicated by higher FSC values for the half map used for refinement compared to the second half map, and worse model geometry (Extended Data Fig. 3d). Conversely, higher weighing of model geometry leads to a poorer fit to the map, reflected in higher R-factors and a lower FSC against the final map, but also reduces overfitting and results in a better model geometry (Extended Data Fig. 3e).

Nomenclature of ribosomal proteins. A new scheme for naming of ribosomal proteins was recently proposed¹⁰ and adopted by the field. We are using this new nomenclature system for naming of mitoribosomal proteins. For all mitoribosomal proteins with homologues in yeast mitoribosomes or in bacterial ribosomes, names have been established^{5,10}. For mitoribosomal proteins without homologues in yeast, we use names that are as similar to the old naming system as possible (for example, MRPL39 is now termed mL39). However, for mitoribosomal proteins MRPL57, MRPL58 (ICT1) and MRPL59 (CRIF1), this was not possible, as the corresponding mLXX names have been assigned to non-homologous yeast proteins⁵. We therefore use mL62 (MRPL58, ICT1), mL63 (MRPL57, MRP63) and mL64 (MRPL59, CRIF1) for these proteins (Supplementary Table 1). Furthermore, as concluded from the BackPhyre⁵⁰ searches mentioned above, MRPL55 is the mitoribosomal homologue of bacterial bL31 and has been designated bL31m.

Creation of figures. Figures depicting molecular structures were created using the UCSF Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081)⁵⁶ and PyMOL (The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.). Local resolution plots were generated in ResMap⁵⁷.

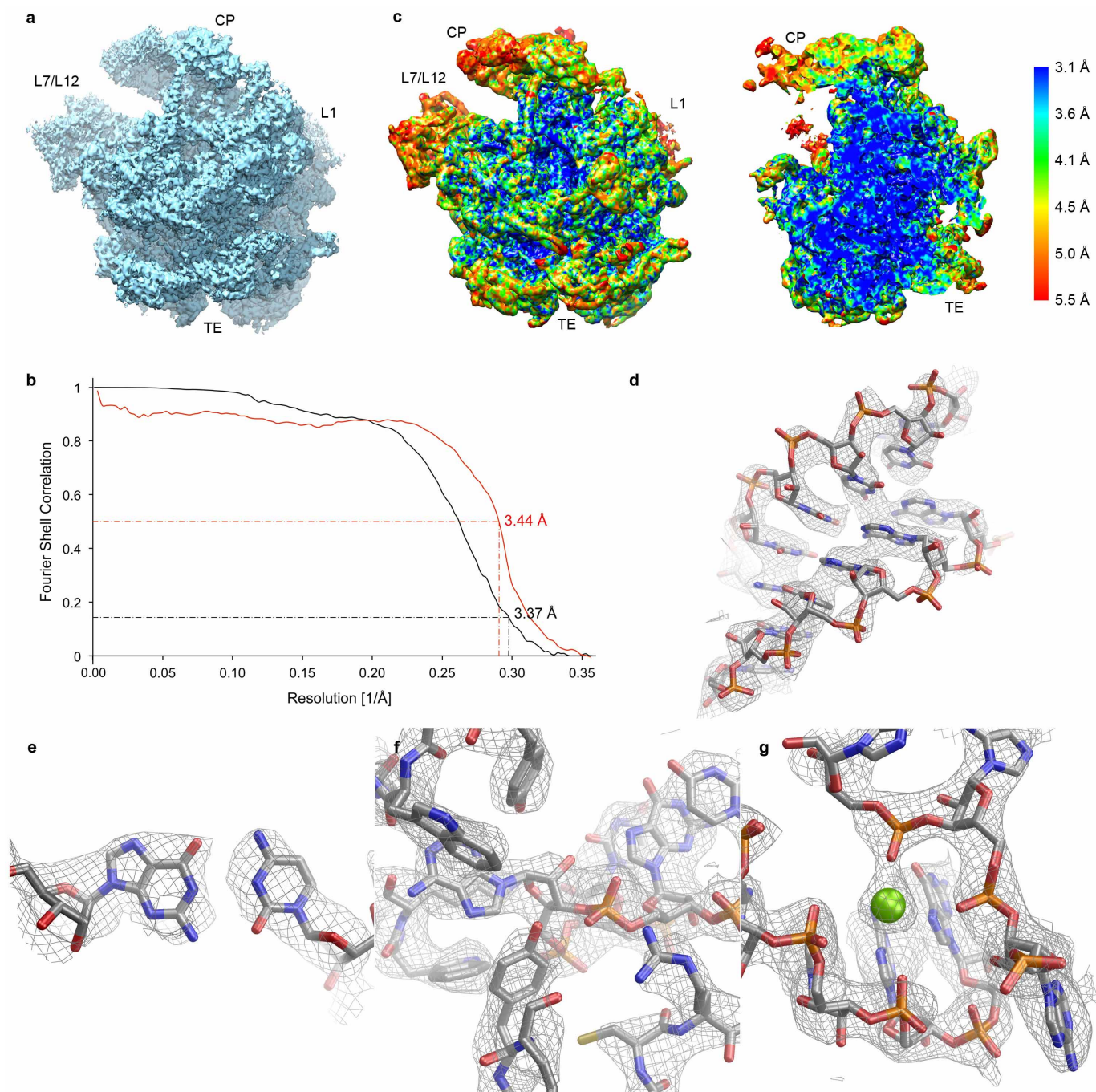
31. Tivol, W. F., Briegel, A. & Jensen, G. J. An improved cryogen for plunge freezing. *Microsc. Microanal.* **14**, 375–379 (2008).
32. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
33. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
34. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
35. Fernández, I. S., Bai, X.-C., Murshudov, G., Scheres, S. H. W. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157**, 823–831 (2014).
36. Leitner, A. *et al.* Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. USA* **111**, 9455–9460 (2014).
37. Leitner, A. *et al.* Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Mol. Cell. Proteom.* **11**, M111.014126 (2012).
38. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nature Protocols* **9**, 120–137 (2014).
39. Walzthoeni, T. *et al.* False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nature Methods* **9**, 901–903 (2012).
40. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
41. Deutsch, E. W. *et al.* A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **10**, 1150–1159 (2010).
42. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
43. Koskinen, V. R., Emery, P. A., Creasy, D. M. & Cottrell, J. S. Hierarchical clustering of shotgun proteomics data. *Mol. Cell. Proteom.* **10**, M110.003822 (2011).
44. Schilling, B. *et al.* Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol. Cell. Proteom.* **11**, 202–214 (2012).
45. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteom.* **5**, 144–156 (2006).
46. Fabre, B. *et al.* Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. *EUPROT* **4**, 82–86 (2014).
47. Jones, T. A. Interactive electron-density map interpretation: from INTER to O. *Acta Crystallogr. D* **60**, 2115–2125 (2004).
48. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
49. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
50. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).
51. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
52. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
53. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
54. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
55. Urzhumtseva, L., Afonine, P. V. & Adams, P. D. Crystallographic model quality at a glance. *Acta Crystallogr. D* **65**, 297–300 (2009).
56. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
57. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
58. Davydov, I. I. *et al.* Evolution of the protein stoichiometry in the L12 stalk of bacterial and organellar ribosomes. *Nature Commun.* **4**, 1387 (2013).



Extended Data Figure 1 | Computational sorting of the 55S mitoribosome data sets.

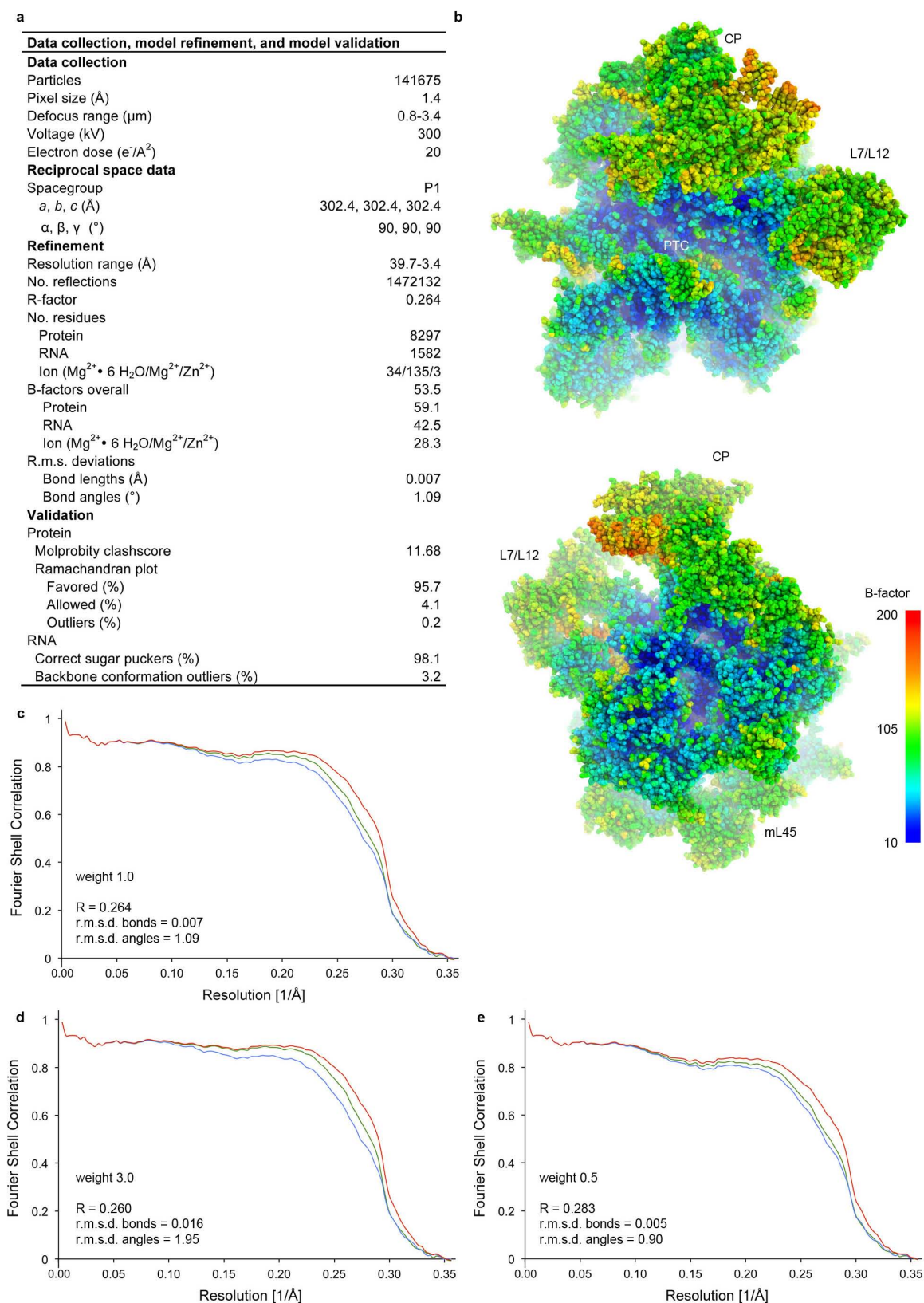
a, To obtain a particle sub-population suitable for high-resolution structure determination, two data sets were computationally sorted in RELION⁹. 2D and 3D classifications (only representative classes are shown) of a data set of a total of roughly 560,000 particle images resulted in the selection of 141,700 55S mitoribosomal particles. The 55S particles were refined to obtain a reconstruction of the 55S mitoribosome (3.6 Å resolution). **b**, The map obtained for the mammalian 55S mitoribosome (28S subunit in pale yellow;

39S subunit in light blue) filtered to 6 Å resolution. Two tRNA molecules are seen in the intersubunit space (A-site tRNA in yellow; P-site tRNA in purple). **c**, Local resolution plot for the 55S reconstruction (surface of the map on the left, viewed in section at the right). The local resolution is clearly below the average of 3.6 Å in some parts of the map, particularly in the 28S subunit (4.1 Å on average for the 28S subunit). Refinement with a mask around the 39S subunit (see **a** and Extended Data Fig. 2) resulted in an improved map suitable for *de novo* model building in the 39S section of the map (3.4 Å resolution).



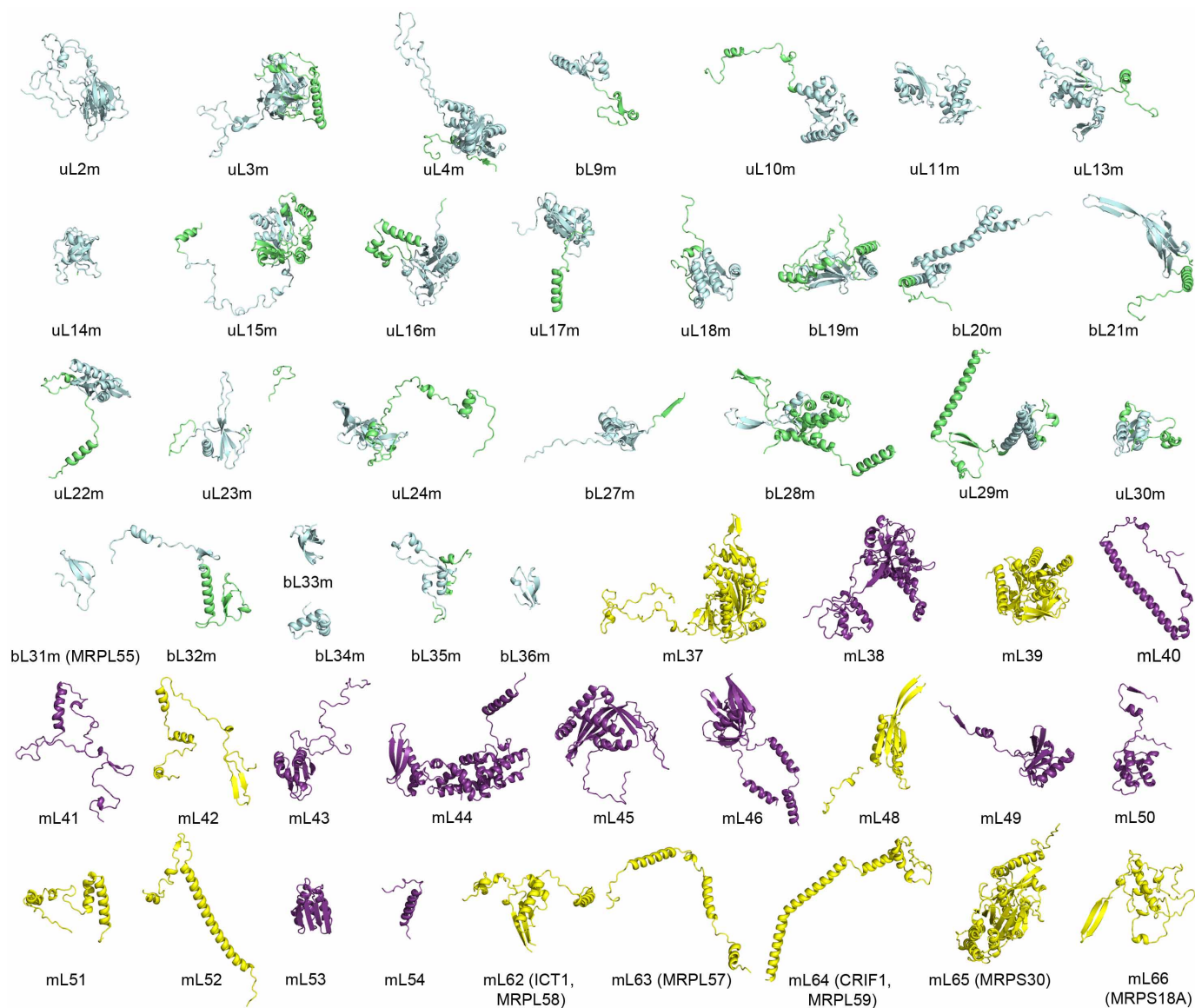
Extended Data Figure 2 | The cryo-EM map of the 39S subunit within the 55S mitoribosome. **a**, Surface rendering of the 39S subunit cryo-EM map at 3.4 Å resolution, viewed from the solvent side. Landmarks are indicated: CP, central protuberance; L1, L1 stalk; L7/L12, L7/L12 stalk base; TE, polypeptide tunnel exit region. **b**, Black: FSC curve of the 39S cryo-EM reconstruction. The resolution is 3.4 Å according to the FSC = 0.143 criterion⁵³. Curve computed using RELION⁹ from auto-masked maps computed from data half-sets.

Red: the FSC curve computed between the final reconstruction and the refined coordinates (see Extended Data Fig. 3) indicates a resolution of 3.4 Å according to the FSC = 0.5 criterion, which has to be used in this case⁵³. **c**, Local resolution plot. **d–g**, Examples for the quality of the density: an rRNA double helix (**d**), a base pair (**e**), interactions of protein and rRNA with many large side chains visible (**f**), and a Mg²⁺ ion coordinated by the rRNA (**g**).



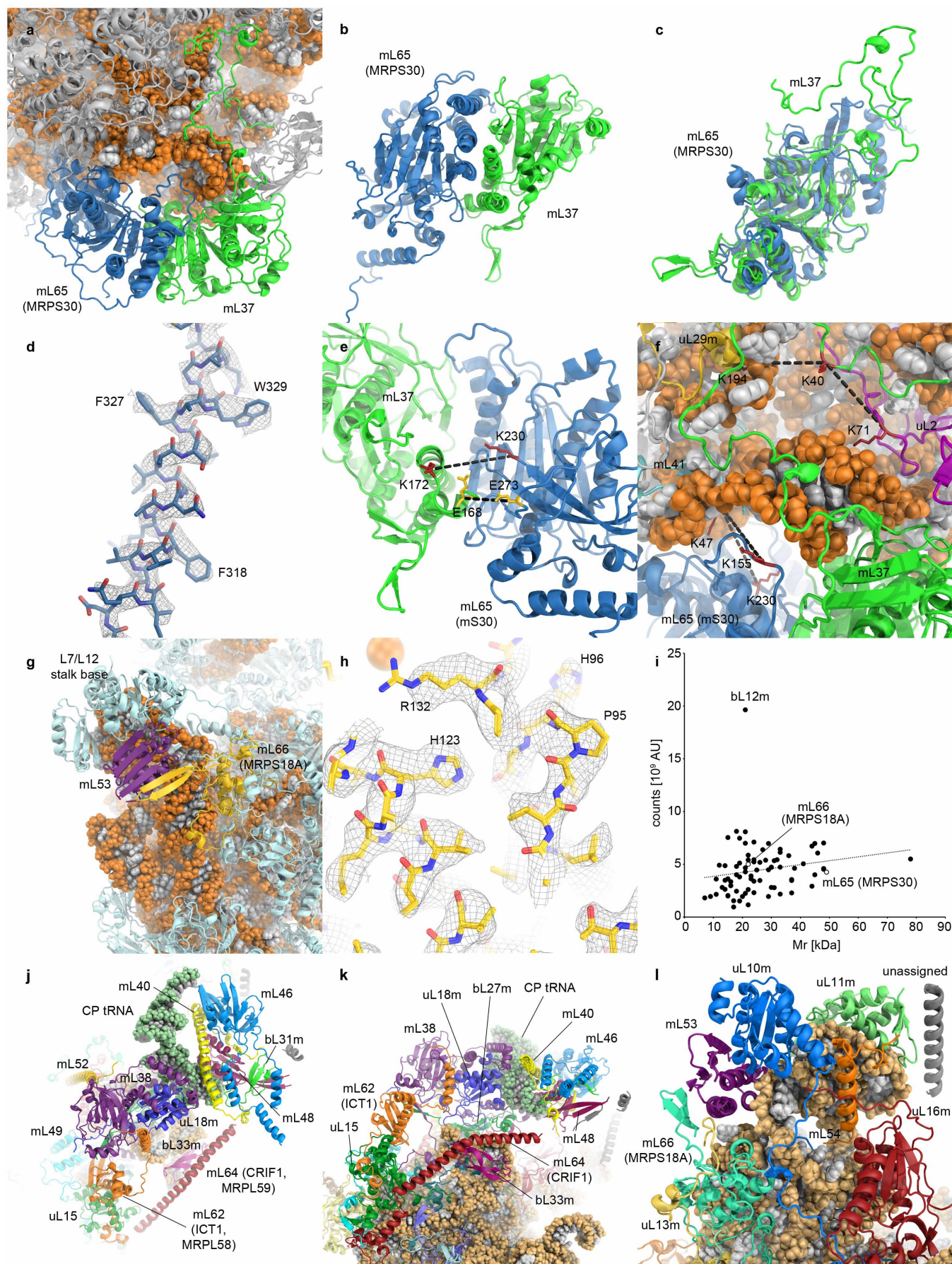
Extended Data Figure 3 | Coordinate refinement of the 39S subunit atomic model. **a**, Model refinement and validation statistics. **b**, Distribution of B-factors in the atomic model of the 39S subunit. Higher B-factors at the periphery, particularly in the L7/L12 stalk and CP regions, reflect the lower local resolution in these areas (Extended Data Fig. 2c). **c**, Refinement weight optimization. FSC curves of the atomic coordinates refined into maps

calculated from one-half of the data (red, FSC model versus map computed from the full data set; blue, FSC model versus half-set 1 (used for test refinement); green, model versus half-set 2 (not used for test refinement)) using a weight of 1.0 for the reciprocal space data (see Methods for details). **d**, As in **c**, but with a weight of 3.0. **e**, As in **c**, but with a weight of 0.5.



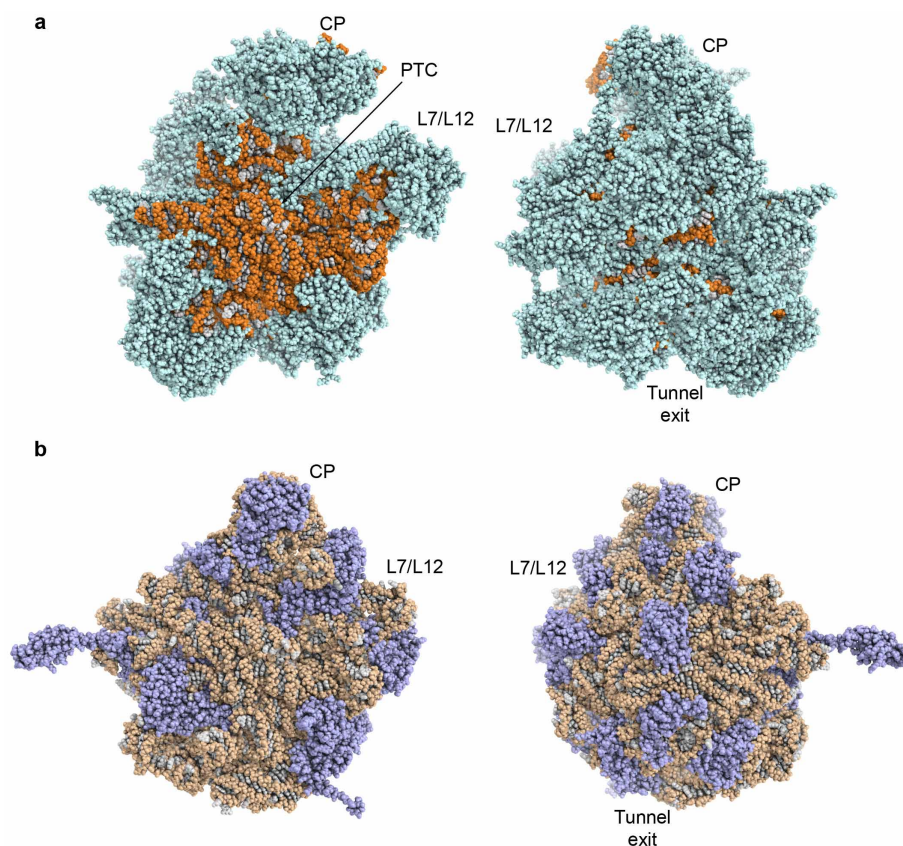
Extended Data Figure 4 | Overview of protein folds in the 39S subunit.
Overview of the structures of the 39S subunit proteins. Cyan, homologues of bacterial ribosomal proteins, with extensions in lime green; purple,

mitochondrial-specific ribosomal proteins that are also present in yeast; yellow, mitoribosomal proteins without homologues in yeast.



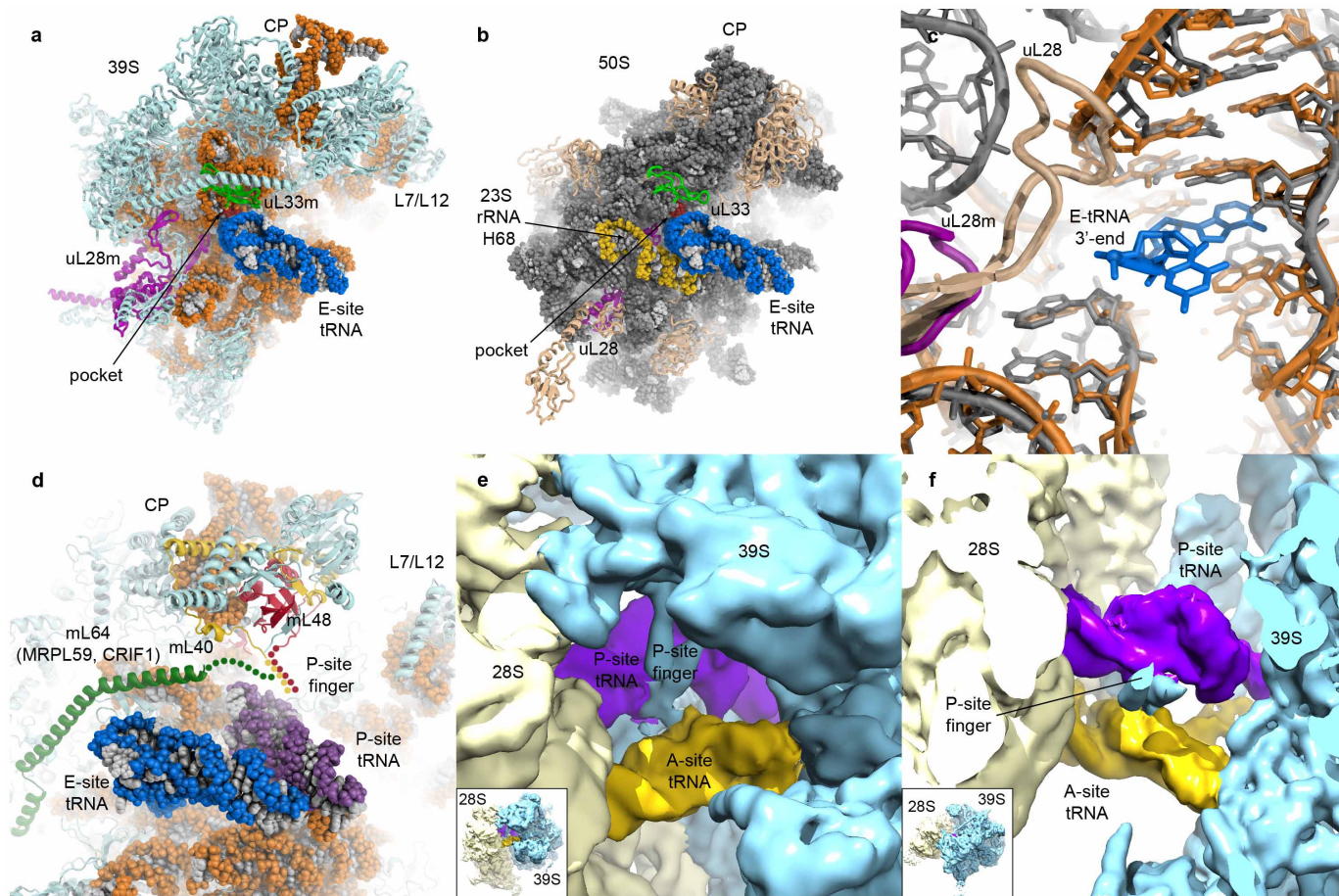
Extended Data Figure 5 | Protein structures in the 39S subunit. **a**, mL37 (green) and mL65 (MRPS30; blue) bound to the 39S subunit (16S rRNA backbone in orange; other mitoribosomal proteins in grey). **b**, mL37 and mL65 (MRPS30) form a pseudo-dimeric assembly on the 39S subunit. **c**, Superposition of mL37 and mL65 (MRPS30), revealing their common core fold with additional extensions specific to each protein. **d**, Segment of the mL65 (MRPS30) structure shown in the cryo-EM map (only density corresponding to the selected structural element is shown). Side chain densities can be clearly identified and allowed the identification of the register of the mL65 (MRPS30) sequence in the map. **e**, Two CX-MS crosslinks between mL65 (MRPS30) and mL37 confirm the placement of mL65 (MRPS30) on the 39S subunit (red, DSS crosslink between lysines; gold, PDH crosslink between carboxyl groups). **f**, Crosslinks with other mitoribosomal proteins. DSS crosslinks of mL37 with uL2m (purple) and uL29m (gold) and of mL65 (MRPS30) with mL41 (cyan) were observed. All crosslink C α -C α distances are 22 Å or shorter. **g**, Binding site of mL66 (MRPS18A) at the L7/L12 stalk base. mL66 (MRPS18A; gold)

and neighbouring mL53 (purple) form a joint β -sheet. **h**, Side-chain density allows the unambiguous assignment of the location and trace of mL66 (MRPS18A). **i**, Semi-quantitative MS analysis of protein abundance in the 55S mitoribosome. mL66 (MRPS18A) and mL65 (MRPS30) are present in roughly stoichiometric amounts with other 39S subunit proteins, indicating that only one copy is present per 55S mitoribosome. Only bL12m was detected with clearly more than one copy per 55S mitoribosome, as expected from the multimeric architecture of the L7/L12 stalk⁵⁸. **j**, Top view of the CP. In addition to the mitochondrial-specific CP tRNA, several mitochondrial-specific proteins have been recruited to the strongly remodelled CP of the 39S subunit. **k**, Side view of the CP. Owing to the missing 5S rRNA, the CP is connected to the subunit main body exclusively by protein contacts. **l**, Structure of the L7/L12 stalk base. Mitochondrial-specific ribosomal proteins mL53, mL54 and mL66 (MRPS18A) have been recruited to the L7/L12 stalk base and may stabilize its connection to the 39S subunit main body.



Extended Data Figure 6 | Overall distribution of rRNA and proteins in the mitochondrial 39S subunit and the bacterial 50S subunit. **a**, Mitochondrial ribosomal large subunit (subunit interface side on the left, solvent-exposed side on the right). Proteins (cyan) almost entirely cover the solvent-exposed side of the 39S subunit, leaving barely any rRNA (orange) exposed. **b**, In bacterial ribosomes¹⁵ (L1 stalk omitted for clarity), a significant portion of the ribosomal surface on the solvent-exposed side is composed of rRNA (brown),

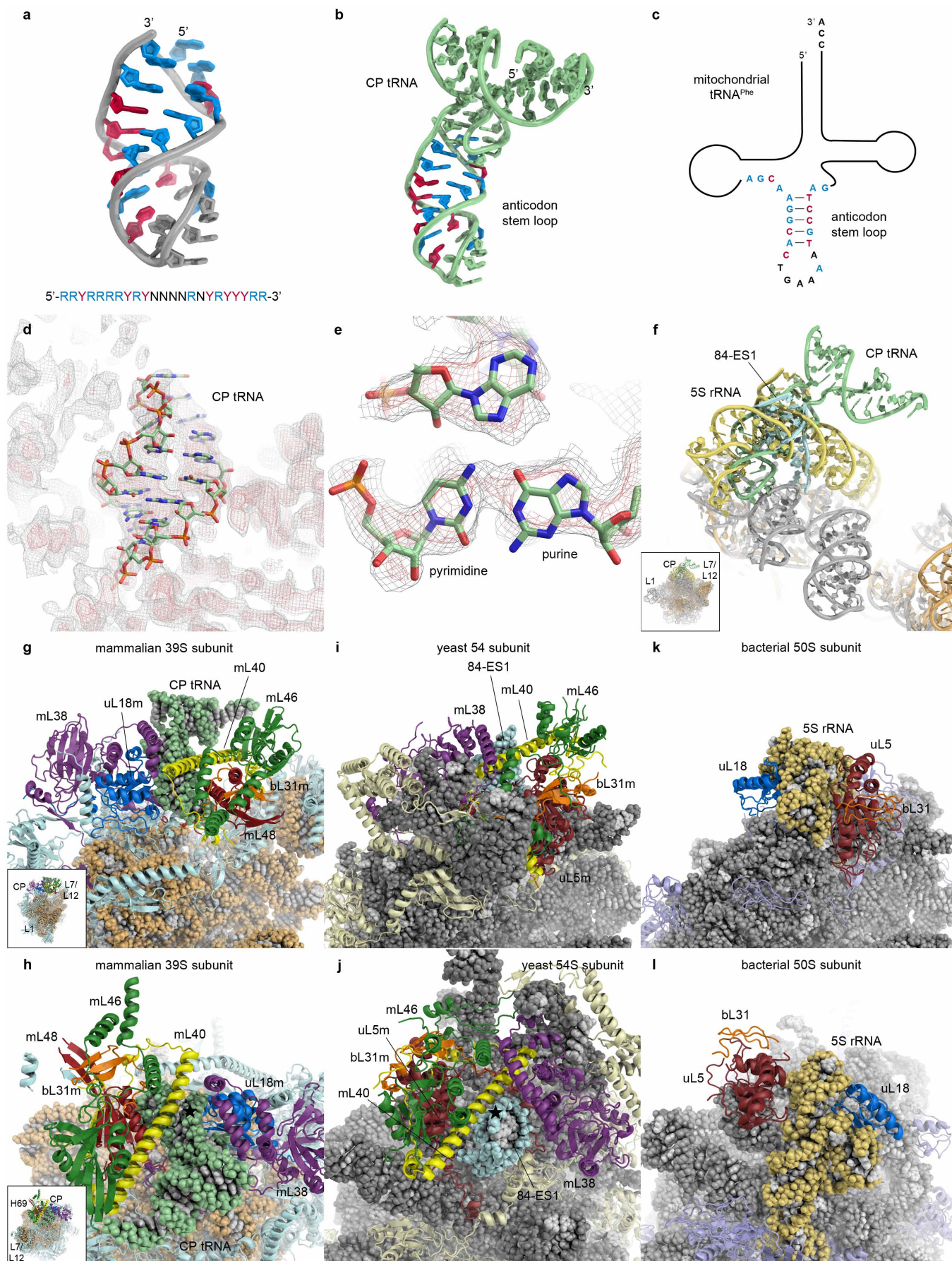
with ribosomal proteins (slate blue) scattered across the subunit surface. The subunit interface side, which makes contact to the small subunit and binds tRNA substrates, is formed by rRNA to a large extent in both types of ribosomes. Landmarks are indicated: L7/L12, L7/L12 stalk base; CP, central protuberance; PTC, peptidyl transferase centre; tunnel exit, polypeptide tunnel exit site.



Extended Data Figure 7 | The E-site and the P-site finger of the 39S subunit.

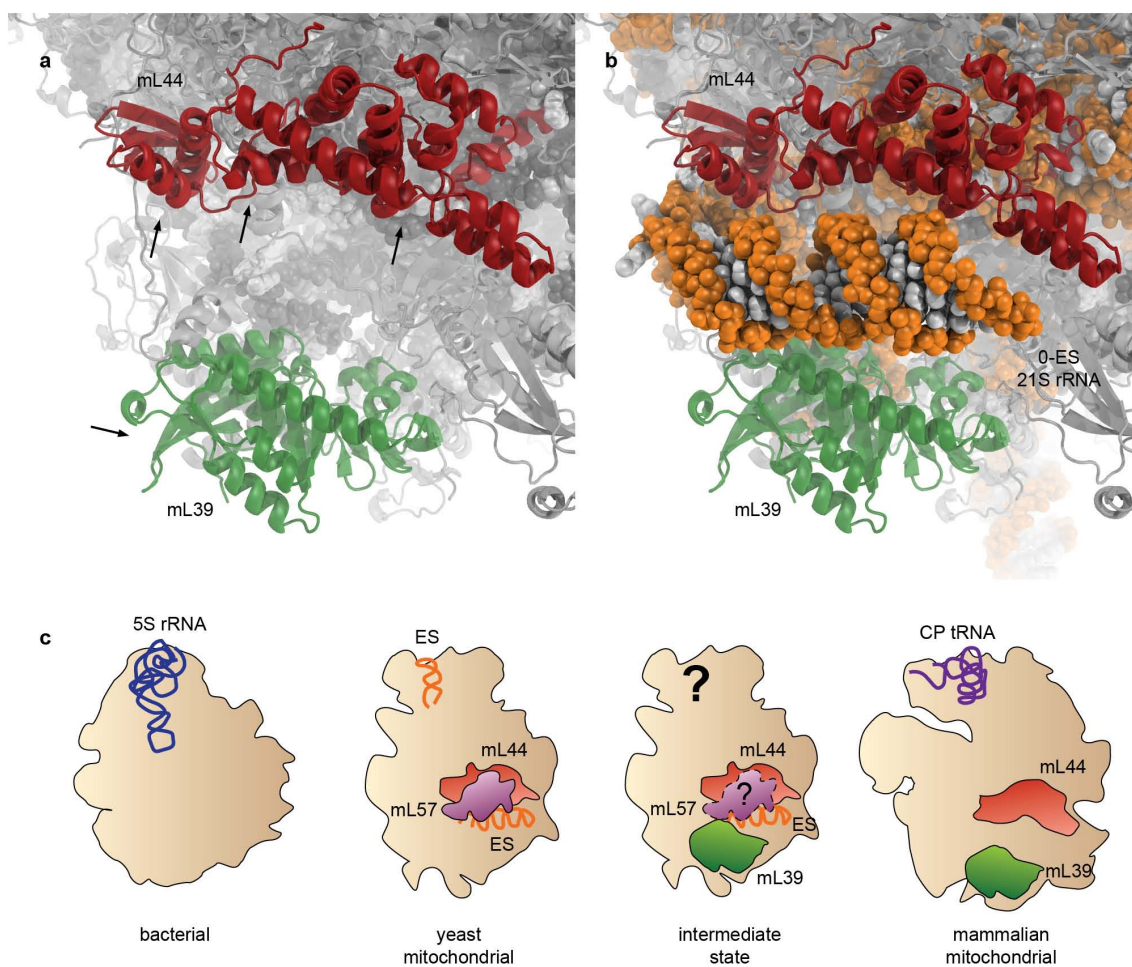
a, 39S subunit (16S rRNA, orange; ribosomal proteins, cyan; uL28m, purple; uL33m, green) with E-site tRNA (blue) modelled based on the superposition of the structure of the bacterial ribosome with bound tRNAs¹⁵. The residues corresponding to the tRNA 3' end binding pocket in bacteria¹⁵ are shown in dark red. **b**, In bacterial ribosomes (23S rRNA, grey; ribosomal proteins, light brown; L1 stalk omitted for clarity, otherwise as in **a**), rRNA helix 68 (yellow) additionally interacts with the E-site tRNA. This structural element is missing in the 39S subunit. **c**, The pocket into which the terminal nucleotide of the E-site tRNA inserts in bacterial 23S rRNA (grey)²⁶ is conserved in the

mitochondrial 16S rRNA (orange). Ribosomal protein uL28 (light brown) interacts with the E-site tRNA in bacteria¹⁵, but is shortened in the 39S subunit (uL28m in purple). **d**, mL40 (yellow), mL64 (CRIF1, MRPL59) (green) and mL48 (red) approach the area where the E- and P-site tRNAs are located based on superposition of the bacterial structure¹⁵. 16S rRNA in orange, 39S proteins in pale cyan. **e**, Cryo-EM density of the 55S mitoribosome (filtered to 6 Å resolution) segmented into 39S and 28S subunits and A- and P-site tRNAs. The P-site finger is inserted between the A- and P-site tRNAs and contacts both tRNAs. **f**, Same as **e**, but shown in top view with part of the ribosome segmented away to reveal the tRNAs.



Extended Data Figure 8 | The architecture of the mammalian and yeast mitoribosomal central protuberances. **a**, The purine–pyrimidine pattern observed in the additional RNA density at the CP as built model (top) and sequence (bottom). Y, pyrimidines; R, purines; N, undetermined. **b**, The model of CP tRNA with bases in the segment corresponding to the pattern in **a** coloured accordingly. **c**, Simplified secondary structure diagram of mitochondrial tRNA^{Phe} with the segment corresponding to the pattern in **a** coloured accordingly. **d**, View of the CP tRNA and the cryo-EM density in the vicinity. Phosphate positions are clearly visible, but the stacked bases merge into continuous strands due to lower local resolution. **e**, Purines and pyrimidines can be distinguished by their shape. **f**, Superposition of the large subunit rRNAs of bacterial¹⁵ (23S rRNA in grey, 5S rRNA in yellow), mammalian mitochondrial (16S rRNA in orange; CP tRNA in light green) and yeast mitochondrial ribosomes⁵ (only 84-ES1 shown for clarity, cyan). **g**, **h**, Side

and top views of the mammalian mitoribosomal CP with the CP tRNA in light green, mL38 in purple, uL18m in blue, mL40 in yellow, mL46 in dark green, mL48 in red and bL31m in orange. **i**, **j**, Side and top views of the yeast mitoribosomal CP⁵ with mL38 in purple, mL40 in yellow, mL46 in dark green, bL31m in orange, uL5m in dark red and 84-ES1 in cyan. Homologous proteins in yeast and mammals occupy roughly similar positions, while mammalian mL48 is positioned similarly to yeast uL5m. Mammalian mL38 is shifted slightly towards the solvent side relative to its yeast counterpart. 84-ES1 occupies a similar position as the anticodon stem-loop part of mammalian CP tRNA, and mL40 binds to the RNA minor groove (star) in both cases. **k**, **l**, View of the bacterial ribosomal CP¹⁵. 23S rRNA in light grey, 5S rRNA in yellow, uL18 in blue, uL5 in dark red, bL31 in orange, and other ribosomal proteins in light blue.



Extended Data Figure 9 | Evolution of RNA binding mitochondrial ribosomal proteins. **a**, In the mammalian mitoribosome, the surfaces of mL39 (green) and mL44 (red) corresponding to the RNA binding surfaces of their RNA binding homologues (arrows) are exposed to the solvent. **b**, The presence of an rRNA expansion segment in the yeast mitoribosome (yeast 21S rRNA⁵ in

orange superimposed on the 39S subunit) explains the arrangement of proteins observed in the mammalian mitoribosome. **c**, Schematic of the evolution of this region of the mitoribosome. Before the loss of the rRNA expansion segments (ESs), mL39, mL44 and mL57 were possibly simultaneously present in evolutionary precursors of mammalian mitoribosomes.

Extended Data Table 1 | CX-MS crosslinks used for confirmation of protein localizations

Peptides	Protein/ residue	Protein/ residue	M _r (Da)	m/z	Id score
DSS					
KPEPPPLDR-TKYTVMPVK-a1-b2	mL37 K40	uL2m K71	2251.223	751.416	31.98
YQINDKPHNQIR-DGKQVQIK-a6-b3	mL65 K230	mL41 K47	2596.369	866.464	29.84
YQINDKPHNQIR-LWHSTEDIPKR-a6-b10	mL65 K230	mL37 K172	3043.556	761.897	27.78
QWPIPWYLNKK-KPEPPPLDR-a10-b1	uL29m K194	mL37 K40	2657.432	665.366	27.06
KRAPVIQDR-DGKQVQIK-a1-b3	mL65 K155	mL41 K47	2153.232	539.316	26.63
PDH					
QYENTIFIGSK-LWHSTEDIPK-a3-b6	mL65 E273	mL37 E168	2675.37	669.85	25.92

Peptides, crosslinked peptides including location of the crosslinked residues in the peptides; protein/residue, protein name and residue number of crosslinked amino acids; M_r, theoretical molecular mass of the crosslink product; m/z, experimentally determined mass-to-charge ratio; Id score, identification score calculated by xQuest.

Structural basis for the assembly of the Sxl–Unr translation regulatory complex

Janosch Hennig^{1,2}, Cristina Militti^{3,4}, Grzegorz M. Popowicz^{1,2}, Iren Wang^{1,2}, Miriam Sonntag^{1,2}, Arie Geerlof¹, Frank Gabel^{5,6,7,8}, Fátima Gebauer^{3,4} & Michael Sattler^{1,2}

Genetic equality between males and females is established by chromosome-wide dosage-compensation mechanisms. In the fruit fly *Drosophila melanogaster*, the dosage-compensation complex promotes twofold hypertranscription of the single male X-chromosome and is silenced in females by inhibition of the translation of *msl2*, which codes for the limiting component of the dosage-compensation complex^{1,2}. The female-specific protein Sex-lethal (Sxl) recruits Upstream-of-N-ras (Unr) to the 3' untranslated region of *msl2* messenger RNA, preventing the engagement of the small ribosomal subunit³. Here we report the 2.8 Å crystal structure, NMR and small-angle X-ray and neutron scattering data of the ternary Sxl–Unr–*msl2* ribonucleoprotein complex featuring unprecedented intertwined interactions of two Sxl RNA recognition motifs, a Unr cold-shock domain and RNA. Cooperative complex formation is associated with a 1,000-fold increase of RNA binding affinity for the Unr cold-shock domain and involves novel ternary interactions, as well as non-canonical RNA contacts by the α 1 helix of Sxl RNA recognition motif 1. Our results suggest that repression of dosage compensation, necessary for female viability, is triggered by specific, cooperative molecular interactions that lock a ribonucleoprotein switch to regulate translation. The structure serves as a paradigm for how a combination of general and widespread RNA binding domains expands the code for specific single-stranded RNA recognition in the regulation of gene expression.

Translational repression of *msl2* mRNA is coordinated by Sxl binding to uridine-rich stretches in both untranslated regions (UTRs): binding to the 3' UTR inhibits the recruitment of the small ribosomal subunit whereas binding to the 5' UTR inhibits the scanning of those subunits that presumably have escaped the 3' UTR-mediated control^{4–7} (Fig. 1a). At the 3' UTR, the recruitment of Unr by Sxl in close spatial proximity is critical for translational repression^{5,6}. The region of Sxl containing residues 122–301 (*Drosophila* RNA binding domain 4, dRBD4; Fig. 1b) shows full translational repression activity, while the RNA recognition motifs (RRMs) alone (residues 123–294, dRBD3) are necessary and sufficient for RNA binding⁷. Only the first cold-shock domain (CSD1)⁸ of Unr is required for complex formation with Sxl and *msl2* mRNA⁹. Notably, CSD1 and Sxl do not interact in the absence of RNA, suggesting a cooperative binding mechanism (Extended Data Fig. 1a). A 46-nucleotide region in the *msl2* 3' UTR containing two uridine-rich Sxl-binding sites (sites E and F, Fig. 1c and Extended Data Fig. 1b) is sufficient for complex formation and translational repression^{5,10}. To identify the minimal region required for Unr and Sxl binding, we analysed ternary complex formation by electrophoretic mobility shift assays (EMSA) (Fig. 1d and Extended Data Fig. 1b) using wild-type and variant RNAs. Binding of dRBD4 and Unr to the wild-type RNA indicated the presence of two complexes (asterisks in Fig. 1d). The number of complexes was reduced to one by mutation of either Sxl-binding site, and site F supported complex formation with a higher affinity than site E (compare mutants 1 and 3 in Fig. 1d). Mutation of the sequences

surrounding site F affected Unr binding, while more distal mutations did not impair complex formation (Fig. 1d and mutants 5 and 6 in Extended Data Fig. 1b). These data indicate the formation of 2:2:1 and 1:1:1 dRBD4–Unr–RNA complexes representing the two bands of slower mobility, which was further confirmed by static light scattering measurements (Extended Data Fig. 1c).

To unravel the structural basis of the Sxl–Unr–RNA interactions we determined the crystal structure of the 1:1:1 complex at 2.8 Å resolution

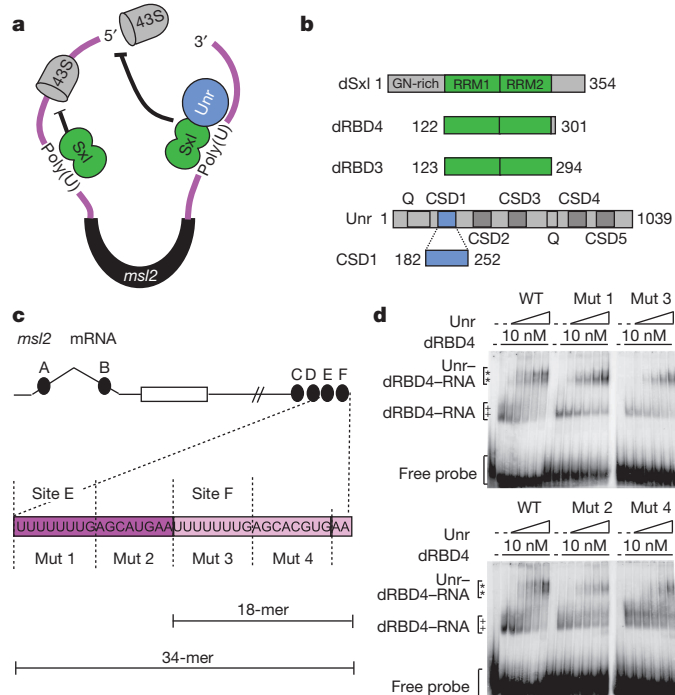


Figure 1 | Analysis of Unr and Sxl binding to *msl2* mRNA. **a**, Model for translational repression of *msl2* mRNA by Sxl and Unr. **b**, Sxl and Unr constructs used in this study. Amino acid numbers are indicated (Q, glutamine-rich domain). **c**, Schematic representation of *msl2* mRNA. Uridine-rich Sxl binding sites are depicted as filled circles (A–F). The first 34 nucleotides of the relevant region of *msl2* 3' UTR required for translational repression are indicated (EF RNA, total length of 46 nucleotides), with two minimal sequences that support Sxl–Unr complex formation (this study) highlighted in dark and light pink. **d**, EMSAs using wild type or mutated versions of EF RNA, where blocks of 8 nucleotides were replaced by CU repeats (see Extended Data Fig. 1 for EMSAs on mutant 5 and mutant 6 RNAs). Amounts of Unr increased for each lane (0.14, 1.4, 4.2, 14 and 70 nM). The positions of the protein–RNA complexes and the free probe are indicated. Two dRBD4–RNA (+) and two Unr–dRBD4–RNA (*) complexes with different stoichiometry are observed. WT, wild type; Mut, mutant.

¹Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, DE-85764, Germany. ²Center for Integrated Protein Science Munich at Biomolecular NMR Spectroscopy, Department Chemie, Technische Universität München, Lichtenbergstr. 4, DE-85747 Garching, Germany. ³Centre for Genomic Regulation, Gene Regulation, Stem Cells and Cancer Programme, Dr Aiguader 88, 08003 Barcelona, Spain. ⁴Universitat Pompeu Fabra, Dr Aiguader 88, 08003 Barcelona, Spain. ⁵Université Grenoble Alpes, Institut de Biologie Structurale, F-38044 Grenoble, France. ⁶Centre National de la Recherche Scientifique, Institut de Biologie Structurale, F-38044 Grenoble, France. ⁷Commissariat à l'Énergie Atomique et aux Énergies Alternatives, Institut de Biologie Structurale, F-38044 Grenoble, France. ⁸Institut Laue–Langevin, F-38042 Grenoble, France.

(Fig. 2 and Extended Data Table 1). In the ternary complex, Sxl binds to the 5' half (containing site F) of the 18-mer UUUUUUUGAGCAC GUGAA RNA, while CSD1 interacts with the 3' flanking sequence. Sxl dRBD3 recognizes the 5' UUUUUUUG sequence (nucleotides 1–8) in a very similar manner as reported for the Sxl complex with *transformer* pre-mRNA¹¹ (protein backbone coordinate root mean square deviation = 0.65 Å; Extended Data Fig. 1d). The fold of CSD1 is similar to other CSDs¹² (Extended Data Fig. 1e). Unexpectedly, the purine-rich sequence downstream of site F (AGCACGUG, nucleotides 9–16) is sandwiched by both proteins. Nucleotides A9 and C11 are specifically recognized by contacts of residues in Sxl and CSD1, thus forming an intertwined network of ternary interactions resembling a 'triple zipper' (Fig. 2b, c). The triple zipper involves Sxl Tyr 164, CSD1 Arg 239, *msl2* C11, CSD1 His 213, Sxl Tyr 160 and *msl2* A9 (Fig. 2b) and is formed by base/aromatic side-chain stacking interactions as well as electrostatic interactions. The C11 base bridges Sxl and CSD1 and is specifically recognized by a network of contacts with the Sxl Lys 161 main-chain carbonyl, CSD1 Asp 237, CSD1 Arg 239 and Sxl Gly 163 (Fig. 2c).

The downstream nucleotides A12 and C13 are recognized by the canonical RNA binding surface of CSD1, consisting of CSD1 Lys 193, His 196,

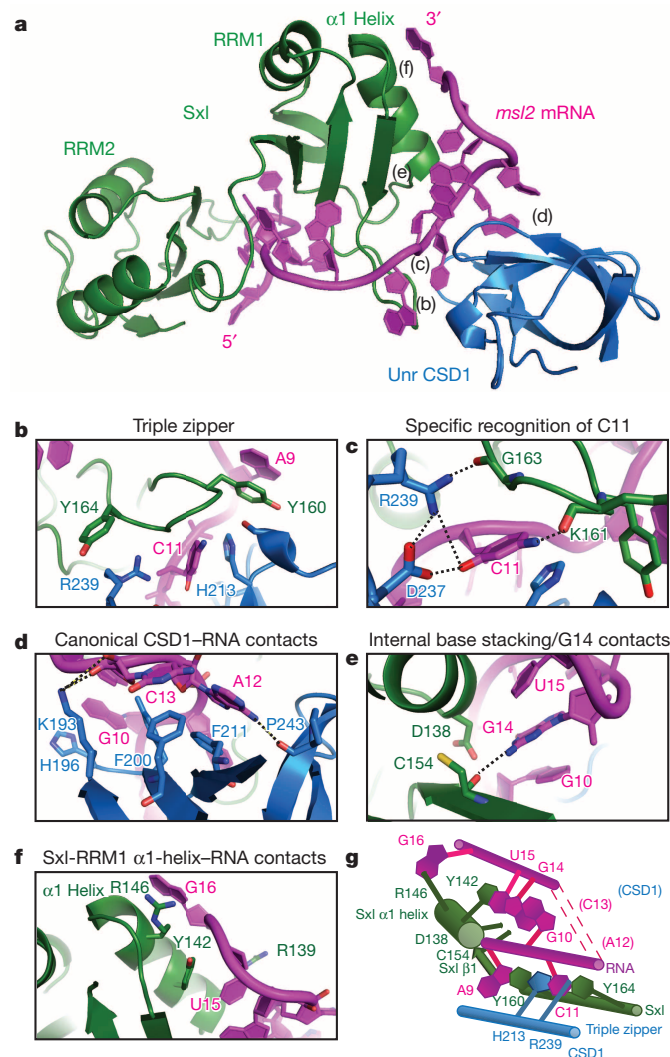


Figure 2 | Crystal structure of the ternary dRBD3-CSD1-*msl2* 18-mer complex at 2.8 Å resolution. a–f, Overview (a) of the ternary complex of Sxl dRBD3 (green), Unr CSD1 (blue) and *msl2* RNA (magenta) and zoomed views (b–f) of specific interactions as indicated. g, Schematic overview of important structural features and binding interfaces in the ternary complex. Sxl β1, β1-strand of Sxl-RRM1.

Phe 200, Phe 211 and Pro 243 (Fig. 2d and Extended Data Fig. 1e)^{12–14}. The CSD1 Lys 193 side chain contacts the 3' phosphate of C13 and thus anchors the phosphodiester backbone for C13–G14. The G14 base stacks with G10 and U15 (G10–G14–U15) and its position is further fixed by contacts with the backbone carbonyl of Sxl Cys 154 (Fig. 2e). These interactions direct the 3' terminal nucleotides (U15 and G16) toward contact with the α1 helix of the Sxl RRM1 domain (Fig. 2f). These non-canonical interactions of U15 (stacking with Sxl Tyr 142 and Arg 139) and G16 (with Sxl Arg 146) are not sequence-specific but depend on the presence of CSD1. Taken together, the *msl2* RNA is sandwiched between Sxl and Unr by a set of unique, intertwined protein–RNA–protein interactions, which mediate cooperative RNA recognition by Sxl RRM1 and Unr CSD1, respectively (Fig. 2g). The crystal structure is further supported by solution NMR and small-angle X-ray and neutron scattering data (Extended Data Fig. 2). NMR chemical shift perturbations for residues in Sxl (Tyr 142, Tyr 146, Tyr 160 and Tyr 164) and CSD1 (His 213 and Arg 239; Fig. 3a) confirm the Sxl-RRM1-α1-helix–RNA contacts and the triple-zipper formation. Small-angle X-ray scattering, static light scattering and NMR experiments were also used to characterize the assembly of Sxl, CSD1 and *msl2* RNA in the 34-mer RNA 2:2:1 complex (Extended Data Fig. 3 and Extended Data Table 2). The data show that no additional quaternary contacts are formed in the 2:2:1 complex and therefore essential features of *msl2* RNA binding by Sxl and Unr CSD1 are described by the 18-mer 1:1:1 complex (Fig. 2).

Next, we used isothermal titration calorimetry to determine the contributions of cooperative RNA binding by Sxl and CSD1 (Extended Data Fig. 4a–e and Extended Data Table 3). Titration of a hexamer RNA representing the region of *msl2* that is recognized by CSD1 in the crystal structure yields a dissociation constant (K_d) of 14 μM for CSD1 (Extended Data Fig. 4a). In the complex the affinity increases ~1,000-fold to a K_d of 15 nM (Extended Data Fig. 4e). For Sxl dRBD3 the affinity increases tenfold, from the reported 120 nM for uridine-rich single-stranded RNA¹⁵ to 15 nM in the ternary complex. Thus, Unr and Sxl bind to *msl2* mRNA with extraordinary cooperativity.

To determine the contributions of protein–RNA contacts seen in the crystal structure we monitored the effect of alanine substitutions of residues specifically involved in ternary complex formation using EMSA experiments and *msl2* translation assays in *Drosophila* embryo extracts (Fig. 3b–d and Extended Data Fig. 5). As expected, mutation of residues involved in ternary contacts did not impair the binary Sxl-site-F–RNA interactions (Extended Data Fig. 5a). Substitution of residues which do not interact with either RNA or CSD1 (Thr 137 and Ile 156) showed no effect on ternary complex formation or translational repression (Thr 137, Fig. 3d and Extended Data Fig. 5c). In contrast, alanine substitutions of single Sxl residues involved in RRM1–α1-helix–RNA contacts (Arg 139, Tyr 142 and Arg 146) showed noticeable effects (Fig. 3b and Extended Data Fig. 5d), and a triple mutation (R139A/Y142A/R146A) severely impaired ternary complex formation and almost abolished *msl2* translation repression (Fig. 3b, d). Mutation of residues in the triple zipper Y160A and Y164A of Sxl impaired ternary complex formation and translational repression (Fig. 3c, d) or virtually abolished it (H213A and R239A of CSD1, Fig. 3c). Replacing the side chains of CSD1 Ser 215 or Gln 216 showed no effect (Fig. 3c), as only backbone atoms of these residues mediate RNA contacts (Fig. 2). While the CSD1(H196A) mutation had no visible effect consistent with minor RNA contacts with G10, CSD1(K193A) showed a significantly decreased binding affinity (Extended Data Fig. 5b), due to removal of the hydrogen bond between the RNA backbone and the NH₂ of the Lys 193 side chain (Fig. 2d). This anchor could be important to maintain the internal RNA base stacking between G10, G14 and U15 and to organize the ternary binding interface (Fig. 2). CSD1(D237A) was previously shown to affect complex formation⁹, consistent with its role in recognizing C11 of *msl2* RNA in the triple-zipper motif (Fig. 2c). The specific recognition of C11 is further indicated by reduced complex formation observed for a C11U variant RNA, or for a natural *msl2* sequence lacking triple-zipper nucleotides (Extended Data Fig. 4g–i).

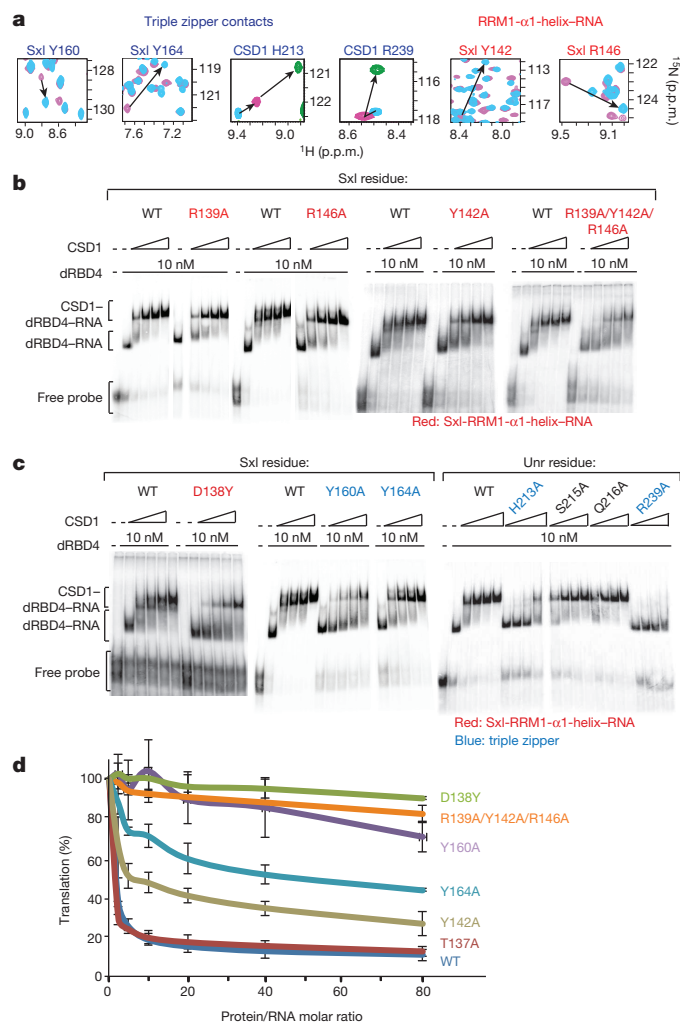


Figure 3 | Ternary complex formation and functional analysis of molecular interfaces in the Sxl-Unr-*msl2*-RNA complex. **a**, NMR chemical shift changes observed upon ternary complex formation for residues that are involved in triple-zipper formation (left two panels: SXL dRBD4 bound to RNA: pink, ternary complex: blue; middle two panels: CSD1 free: blue, bound to RNA: pink, ternary complex: green) or non-canonical Sxl-RNA contacts (two right-most panels; Sxl dRBD4 bound to RNA, pink; Sxl ternary complex, blue). **b**, Binding of wild type and mutant Sxl dRBD4 proteins to EF RNA in the presence of increasing amounts of CSD1 (10, 30 100 and 500 nM) to test residues involved in non-canonical RNA contacts (red). **c**, Binding of wild-type and mutant Sxl dRBD4 and CSD1 proteins to EF RNA to test residues involved in triple-zipper formation (blue). The D138Y mutation present in *Sxl*^{f7/M1} flies was also tested. Concentrations of CSD1 are as in **b**. **d**, Translational repression assays in embryo extracts (containing endogenous Unr) adding increasing amounts of wild-type or mutant Sxl dRBD4. A luciferase reporter containing the *msl2* EF region in the 3' UTR was used. Error bars represent the standard deviation of three experiments.

The functional significance of the unusual contacts seen in the ternary complex is demonstrated *in vivo* by transfecting male Schneider 2 cells with wild-type Sxl or Sxl variants. Whereas wild-type Sxl decreased the expression of endogenous *msl2*, defective mutants failed to do so (Extended Data Fig. 5e). Notably, the activity observed for the mutants *in vivo* correlates well with their RNA binding and translational repression capacities *in vitro*. The biological relevance of the molecular contacts described here was confirmed by the *Sxl*^{f7/M1} allele^{16–18}. *Sxl*^{f7/M1} flies carry two mutations in the *Sxl* gene: the M1 mutation is a transposon insertion that leads to constitutive Sxl expression and male lethality due to *msl2* repression and inactivation of dosage compensation; the f7 mutation rescues M1-induced lethality. Interestingly, the f7 mutation

corresponds to a D138Y replacement. This mutation is expected to affect ternary Sxl-Unr-*msl2*-mRNA complex formation, as the bulky tyrosine sterically interferes with contacts between Sxl Cys 154 and *msl2* G14, and thus destabilizes the internal base stacking with G10 and U15 (Fig. 2e). The Sxl(D138Y) mutant shows strongly reduced ternary complex formation and impaired *msl2* translational repression (Fig. 3c, d), providing a molecular rationale for the *in vivo* phenotype of *Sxl*^{f7/M1}.

Taken together our data demonstrate that the triple zipper and the non-canonical RNA contacts mediated by Sxl RRM1 are critical for translational regulation by Sxl and Unr. It is important to note that these interactions are essential for *msl2* translational repression, but are dispensable for the regulation of *transformer* pre-mRNA splicing¹⁹, as recognition of the uridine-rich 5' region of *msl2* RNA by Sxl dRBD3 in the ternary complex is virtually identical to that previously observed for *transformer* pre-mRNA¹¹. Therefore, recognition of a uridine-rich RNA sequence by Sxl can play distinct roles in regulating splicing and translation depending on the binding of Unr in close proximity.

Our data also explain why human Unr can form a complex with *Drosophila* Sxl and *msl2* RNA⁹ as all residues involved are conserved (His 213, Asp 237 and Arg 239, Extended Data Fig. 6a). On the other hand, CSD1 alone can bind a variety of distinct RNA sequences with similar affinity in the absence of Sxl (Extended Data Fig. 4f). This indicates that strong and specific RNA recognition for the GCACG motif in *msl2* RNA depends on the presence of Sxl in the ternary complex. Interestingly, C11 does not conform to the previously reported consensus sequence²⁰ for human Unr CSD1, but is nevertheless strictly conserved in the *msl2* mRNA of organisms that may employ *D. melanogaster*-like dosage compensation^{21,22} (Extended Data Fig. 7). Consistent with this, CSD1 Asp 237 and Arg 239 (Fig. 2), which recognize C11, are conserved in CSD1 but not in CSD2–5 of Unr proteins⁹ or in cold shock domains of other proteins^{12–14} (Extended Data Fig. 6b).

Although the *Drosophila* dosage-compensation mechanism is not conserved in mammals, we expect that ternary interactions involving RRM and CSD domains with RNA may be important for other biological functions. For example, human orthologues of the proteins studied here, such as the Sxl orthologue HuR or the RNA binding protein RBM6, share triple-zipper and α 1-helix residues (Extended Data Fig. 6c), which could mediate similar interactions.

Sandwiching of single-stranded RNA by multiple proteins has been observed previously, for example in small nuclear ribonucleic particles^{23,24} or the exon junction complex²⁵, but the intertwined recognition observed here is particularly intriguing. Moreover, the combination of these two general and abundant RNA binding domains (RRM and CSD), which are also involved in other RNA binding events, generates a new and unique binding specificity for single-stranded RNA. The intertwined cooperative binding of Sxl and Unr establishes a functionally active arrangement of multiple RNA binding domains from two distinct proteins, thus extending principles recently observed for multi-domain RNA binding proteins^{26,27}.

Our results show that repression of a biological process with dramatic consequences for viability depends on the establishment of a specific set of novel molecular interactions. This is of particular significance considering that a limited set of RNA binding modules has been identified in the mRNA interactome^{28,29}. The Unr-Sxl-*msl2* complex illustrates how the combinatorial activity of general RNA binding domains expands the code for RNA recognition by establishing unique and distinct ribonucleoprotein architectures and thus greatly amplifying the opportunities for regulation of gene expression.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 January; accepted 16 July 2014.

Published online 7 September 2014.

- Conrad, T. & Akhtar, A. Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nature Rev. Genet.* **13**, 123–134 (2012).

2. Gelbart, M. E. & Kuroda, M. I. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development* **136**, 1399–1410 (2009).
3. Graindorge, A., Militti, C. & Gebauer, F. Posttranscriptional control of X-chromosome dosage compensation. *Wiley Interdiscip. Rev. RNA* **2**, 534–545 (2011).
4. Beckmann, K., Grskovic, M., Gebauer, F. & Hentze, M. W. A dual inhibitory mechanism restricts *msl-2* mRNA translation for dosage compensation in *Drosophila*. *Cell* **122**, 529–540 (2005).
5. Abaza, I., Coll, O., Patalano, S. & Gebauer, F. *Drosophila* UNR is required for translational repression of *male-specific lethal 2* mRNA during regulation of X-chromosome dosage compensation. *Genes Dev.* **20**, 380–389 (2006).
6. Duncan, K. *et al.* Sex-lethal imparts a sex-specific function to UNR by recruiting it to the *msl-2* mRNA 3' UTR: translational repression for dosage compensation. *Genes Dev.* **20**, 368–379 (2006).
7. Grskovic, M., Hentze, M. W. & Gebauer, F. A co-repressor assembly nucleated by Sex-lethal in the 3'UTR mediates translational control of *Drosophila msl-2* mRNA. *EMBO J.* **22**, 5571–5581 (2003).
8. Schindelin, H., Marahiel, M. A. & Heinemann, U. Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature* **364**, 164–168 (1993).
9. Abaza, I. & Gebauer, F. Functional domains of *Drosophila* UNR in translational control. *RNA* **14**, 482–490 (2008).
10. Gebauer, F., Grskovic, M. & Hentze, M. W. *Drosophila* sex-lethal inhibits the stable association of the 40S ribosomal subunit with *msl-2* mRNA. *Mol. Cell* **11**, 1397–1404 (2003).
11. Handa, N. *et al.* Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–585 (1999).
12. Sachs, R., Max, K. E., Heinemann, U. & Balbach, J. RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. *RNA* **18**, 65–76 (2012).
13. Mayr, F., Schutz, A., Doge, N. & Heinemann, U. The Lin28 cold-shock domain remodels *pre-let-7* microRNA. *Nucleic Acids Res.* **40**, 7492–7506 (2012).
14. Nam, Y., Chen, C., Gregory, R. I., Chou, J. J. & Sliz, P. Molecular basis for interaction of *let-7* microRNAs with Lin28. *Cell* **147**, 1080–1091 (2011).
15. McLaughlin, K. J., Jenkins, J. L. & Kielkopf, C. L. Large favorable enthalpy changes drive specific RNA recognition by RNA recognition motif proteins. *Biochemistry* **50**, 1429–1431 (2011).
16. Cline, T. W. Autoregulatory functioning of a *Drosophila* gene product that establishes and maintains the sexually determined state. *Genetics* **107**, 231–277 (1984).
17. Bernstein, M., Lersch, R. A., Subrahmanyam, L. & Cline, T. W. Transposon insertions causing constitutive Sex-lethal activity in *Drosophila melanogaster* affect Sxl sex-specific transcript splicing. *Genetics* **139**, 631–648 (1995).
18. Maine, E. M., Salz, H. K., Cline, T. W. & Schedl, P. The Sex-lethal gene of *Drosophila*: DNA alterations associated with sex-specific lethal mutations. *Cell* **43**, 521–529 (1985).
19. Valcárcel, J., Singh, R., Zamore, P. D. & Green, M. R. The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* **362**, 171–175 (1993).
20. Triqueneaux, G., Velten, M., Franzon, P., Dautry, F. & Jacquemin-Sablon, H. RNA binding specificity of Unr, a protein with five cold shock domains. *Nucleic Acids Res.* **27**, 1926–1934 (1999).
21. *Drosophila* 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
22. Park, S. W. *et al.* An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys16 acetylation on the *Drosophila* X chromosome. *Genetics* **177**, 1429–1437 (2007).
23. Weber, G., Trowitzsch, S., Kastner, B., Luhrmann, R. & Wahl, M. C. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J.* **29**, 4172–4184 (2010).
24. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475–480 (2009).
25. Bono, F., Ebert, J., Lorentzen, E. & Conti, E. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**, 713–725 (2006).
26. Mackereth, C. D. *et al.* Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475**, 408–411 (2011).
27. Mackereth, C. D. & Sattler, M. Dynamics in multi-domain protein recognition of RNA. *Curr. Opin. Struct. Biol.* **22**, 287–296 (2012).
28. Baltz, A. G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
29. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).

Acknowledgements We thank H.-S. Kang, M. Pabis and L. Warner for discussions and reading the manuscript. We acknowledge the ILL, Grenoble for BAG SANS beamtime on D22 (local contact A. Martel), the ESRF for beamtime on BM29 (local contact L. Zerrad), the Swiss Light Source for beamtime, the crystallization facility at the Max-Planck-Institute for Biochemistry, Martinsried, and the Bavarian NMR Center for NMR measurement time. We thank T. Madl and C. Göbl for assisting in SAXS measurements of the 34-mer complex. J.H. acknowledges postdoctoral fellowships from the Swedish Research Council (Vetenskapsrådet) and the European Molecular Biology Organization (EMBO, ALTF276-2010). C.M. acknowledges a La Caixa Foundation fellowship. Funding support is acknowledged by grants BFU2009-08243 and Consolider CSD2009-00080 from the Spanish Ministry of Economy and Competitiveness (to F.Ge.) and the Deutsche Forschungsgemeinschaft, grants SFB1035 and GRK1721 (to M.Sa.).

Author Contributions J.H. produced samples, performed NMR, SAXS/SANS, biophysical measurements, crystallization and model building/refinement. C.M. produced samples, performed and analysed EMSA, transfection and translation assays. G.M.P. acquired diffraction data and performed molecular replacement and model building. I.W., M.Sa. and A.G. contributed to sample production. F.Ga. performed and analysed SANS measurements. J.H., F.Ge., C.M. and M.Sa. interpreted results and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The atomic coordinates have been deposited in the Protein Data Bank under accession number 4QQB. The NMR chemical shifts are deposited in the Biological Magnetic Resonance Data Bank, entries 25059, 25060 and 25078 for chemical shifts of CSD1 in the free, RNA-bound and SXL-RNA-bound state, respectively, and 25072 for SXL chemical shifts in the ternary complex. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Sa. (sattler@helmholtz-muenchen.de) or F.Ge. (fatima.gebauer@crg.eu).

METHODS

Plasmids, protein expression, purification and sample preparation. Plasmids for Sxl dRBD3 and Unr CSD1 (pTRX-Sxl, pTRX-CSD1; UniProt: P19339 and Q9VSK3, respectively) were derived from pET-24d(+) (Novagen/Merck), and comprise a His₆-tag and a thioredoxin solubility tag connected via a tobacco etch virus protease (TEV)-cleavage site to Sxl residues 123–294 and Unr residues 185–252, respectively (GAMA remains N-terminally for both proteins after TEV cleavage). Expression and purification of proteins were performed as previously described³⁰. Proteins for crystallization, isothermal titration calorimetry, static light scattering, and small-angle X-ray scattering (SAXS) were expressed in LB medium, whereas proteins for NMR spectroscopy and small-angle neutron scattering (SANS) were expressed in differently isotope labelled M9 minimal medium (supplemented with ¹⁵NH₄Cl and/or ¹³C and/or ²H glucose and/or M9 salts dissolved in 100% D₂O (all isotopes were purchased from Sigma)). The two proteins were then mixed in a 1:1 ratio before addition of 1.1 equivalents of RNA oligonucleotides (synthesized, IBA GmbH, sequences derived from *msl2* mRNA, UniProt: P05034) and subsequent size-exclusion chromatography on a S75 gel-filtration column and concentration to desired values in centrifugation concentrators (3.5 kDa cut-off). For SANS measurements, the buffer was exchanged to NMR buffer with different D₂O concentrations (0, 42% and 70%). E/F site, dRBD4, Unr and CSD1 constructs were previously described^{5,7,10}. dRBD4 and CSD1 mutant derivatives were obtained using the Quick-site directed mutagenesis system (Agilent). Recombinant proteins were expressed and purified from *Escherichia coli*. dRBD4 and derivatives were expressed as glutathione S-transferase fusions and purified as described⁷. His-tagged, full-length Unr was purified following the pET system user manual (Novagen). CSD1 and mutant derivatives were expressed as maltose-binding protein fusions and purified following the pMALc system user manual (New England Biolabs). All proteins were dialysed against buffer D (20 mM HEPES pH 8.0, 20% glycerol, 1 mM DTT, 0.01% NP-40, 0.2 mM EDTA). Protein quality was assessed by Coomassie staining.

Static light scattering. Static light scattering was measured with a Malvern Viscotek instrument (TDA 305) connected downstream to an Äkta purifier equipped with an analytical size-exclusion column (Superdex 75 10/300 GL, GE Healthcare). Concentrations varied for each sample between 2–4 mg ml⁻¹ dependent on the size of the complex and 100 µl of sample was injected. As running buffer, NMR buffer has been used (10 mM KP, 50 mM NaCl, 10 mM DTT, pH 6). Elution profiles were collected for 30 min with a flow rate of 1 ml min⁻¹ and data were collected using absorbance UV detection at 280 nm, right-angle light scattering (RALS) and refractive index (RI). The molecular weights of separated elution peaks were calculated using OmniSEC software (Malvern). As calibration standard, 4 mg ml⁻¹ bovine serum albumin was used before all experiments and the change in refractive index with respect to concentration was set to 0.185 ml g⁻¹ (ref. 31).

Crystal structure determination. The ternary dSXL_{123–294}-CSD1_{181–252}-*msl2*-mRNA 18-mer (UUUUUUUGAGCAGUGAA) was concentrated to ~15 mg ml⁻¹ and crystallized in 0.1 M LiSO₄ and 1% PEG 3350 at 4 °C. Crystals grew to about 0.25 × 0.1 × 0.1 mm in size with smooth appearance and no apparent macroscopic defects. Crystals were collected using 30% Glycerol as cryoprotectant and diffracted at 2.8 Å resolution and a complete data set was collected at the Swiss Light Source synchrotron. Initial phases were obtained by molecular replacement using the previously published crystal structure of dSxl (PDB: 1B7F) and the programs Phaser³² and the CCP4 program suite³³. The initial R-factor was 0.487. The model was completed by manual model building of CSD1 into the electron density and was refined by cyclic rounds of model building using COOT³⁴ and maximum-likelihood refinement using REFMAC³⁵. The unit cell contains two ternary complexes. The crystal had unusually high solvent content of 77%. Data statistics and quality of the model are summarized in Extended Data Table 1.

Probe synthesis and gel mobility shift assays. Labelled E- and F-site RNA and mutant derivatives were synthesized by *in vitro* transcription using hybridized oligonucleotides containing the T7 promoter upstream of the E- or F-site sequence as templates. For the EMSA analysis involving mutant dRBD4 and CSD1 proteins, wild-type E- and F-site-containing RNA was synthesized using pBSK-EF plasmid as a template⁷. EMSAs were performed as previously described⁹.

In vitro translation assays. *In vitro* translation reactions in *Drosophila* embryo extracts were performed in a final volume of 10 µl as previously described³⁶. Reporter firefly *msl2* mRNA (BmutLEF)¹⁰ was used at a final concentration of 1.7 ng µl⁻¹. Renilla luciferase mRNA was co-translated as internal control. Luciferase values were obtained using the dual luciferase assay system (Promega). Firefly luciferase was corrected for Renilla expression and expressed relative to the activity obtained in the absence of Sxl.

Expression of Sxl derivatives in Schneider 2 cells and Msl2 western blot analysis. Schneider 2 (S2) cells were co-transfected with pAC plasmids expressing green fluorescent protein and wild-type or mutant Sxl proteins using the Effectene Transfection Reagent kit (QIAGEN). Sxl-transfected green-fluorescent-protein-positive cells were sorted by flow cytometry and total protein extracts were obtained by

incubating cell pellets in RIPA buffer (150 mM NaCl, 10 mM TRIS pH 7.5, 0.1% SDS, 1% deoxycholate, 5 mM EDTA, 1% Triton X-100, 1x protease inhibitor cocktail from Roche) for 20 min on ice. Western blots were performed with anti-Msl2, anti-dRBD4 and anti-α-tubulin antibodies at dilutions of 1:1,000, 1:1,000 and 1:2,000, respectively. Western blot signals were quantified using ImageJ.

NMR spectroscopy. Backbone resonance assignments for free dRBD3 and dRBD3 bound to RNA were taken from the Biological Magnetic Resonance Data Bank (accession codes 4029 and 4028, respectively), and confirmed by 3D HNCACB, CBCACONH and HNCA triple resonance experiments³⁷. Binding of dRBD3 to RNA is in the slow exchange regime on the NMR chemical shift time scale. Therefore, RNA-bound dRBD3 to U9 RNA was formed by adding RNA in a 1:1 ratio. Backbone chemical shift assignments for CSD1 were obtained from 3D HNCA, HNCACB and CBCACONH experiments recorded for the free and RNA-bound form. Resonance assignments of dRBD3 and CSD1 residues in the context of the ternary complex were determined using TROSY HNCA experiments³⁸ using ²H, ¹⁵N, ¹³C-labelled protein, whereas the respective binding partner was deuterated only. All NMR resonance assignment experiments were carried out at 298 K on Bruker Avance III spectrometers at field strengths corresponding to 750 and 800 MHz proton Larmor frequency, equipped with TXI room temperature probe and TCI cryo-probe heads, respectively. Spectra were processed with NMRPipe³⁹ and analysed with CARRA (<http://cara.nmr.ch>) and SPARKY⁴⁰. ¹⁵N-¹³C' and ¹H-¹⁵N residual dipolar couplings were recorded using an IPAP-HSQC⁴¹ and HNCO-based NMR experiments⁴² with a 200 µM sample that was aligned in a medium containing 10 mg ml⁻¹ Pf1 phage (Profos AG) as described⁴³. Residual dipolar couplings were analysed and the alignment tensor calculated using Module2 (ref. 44).

Small-angle X-ray scattering. Thirty microlitres of dRBD3-CSD1 18-mer complex sample and buffer were measured at 25 °C at the BioSAXS beamline BM29 at the European Synchrotron Radiation Facility using a 2D Pilatus detector. Ten frames with 2-s exposure time per frame were recorded for each complex and buffer sample, using an X-ray wavelength of λ = 1.008 Å. Measurements were performed in flow mode where samples were pushed through a capillary at a constant flow rate to minimize radiation damage. Frames showing radiation damage were removed before data analysis.

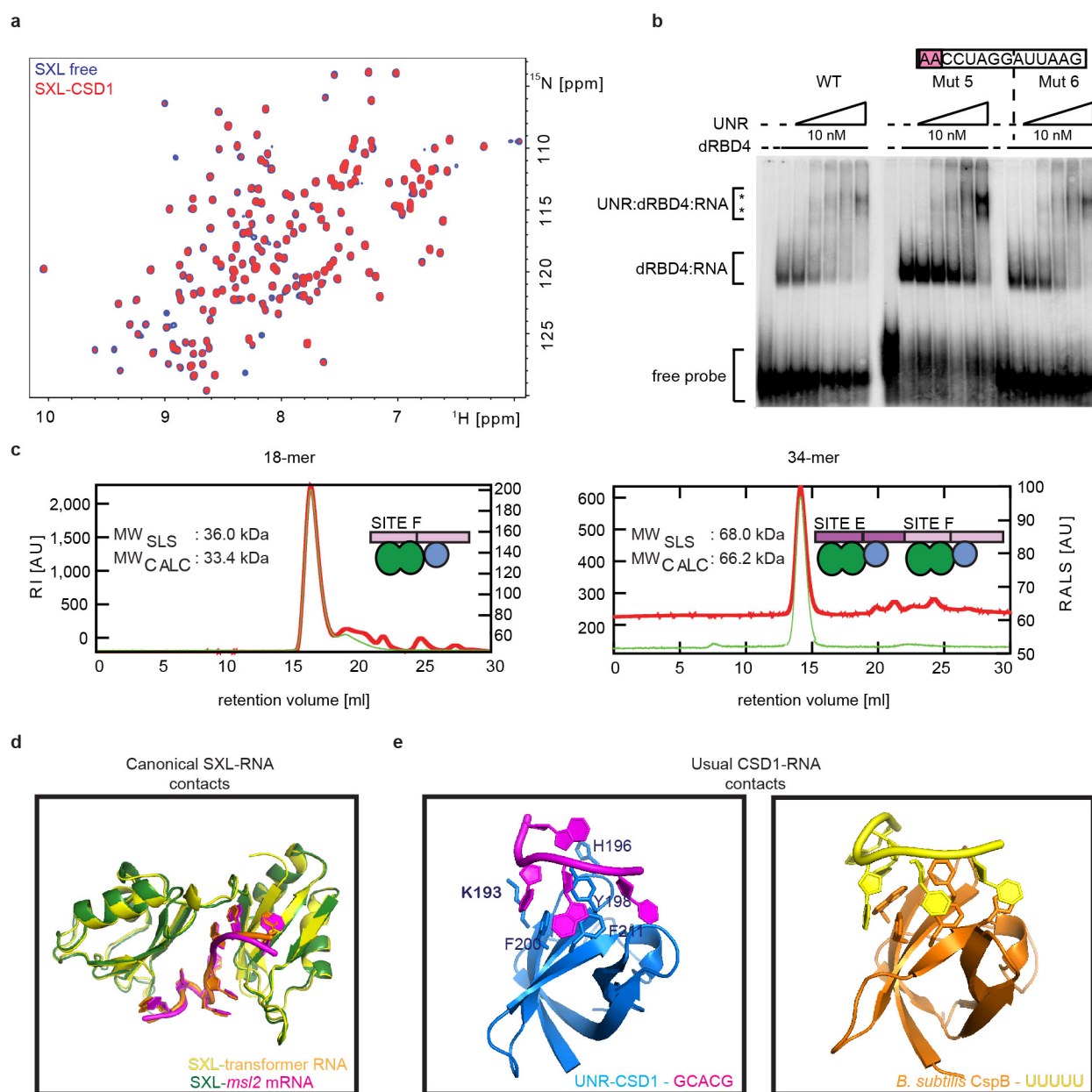
Small-angle neutron scattering. Two hundred microlitres of all samples (including buffers and water; Extended Data Fig. 2 and Extended Data Table 2) were measured in Hellma quartz cells 100QS with 1 mm optical path length at the large dynamic range diffractometer D22 at the Institut Laue-Langevin. Scattering data from all samples were recorded at a 2 m/2 m instrumental detector/collimator configuration (centred detector) at a neutron wavelength λ = 6 Å. H₂O/D₂O buffers, an H₂O water reference sample, the empty beam, an empty quartz cell, as well as a boron sample (electronic background) were measured for data reduction purposes. Exposure times varied between 15 min (boron) and 45 min (empty cell, H₂O and all buffers and samples). Transmissions were measured for 1 min for each sample and the empty beam. All measurements were carried out at 25 °C.

SANS/SAXS data reduction and integral parameters. The raw SANS data were reduced (accounting for detector efficiency, electronic background and angular averaging) using a standard Institut Laue-Langevin software package⁴⁵. For SAXS data collection and processing, the dedicated beamline software BsxCuBE was used in an automated fashion. The one-dimensional scattering intensities of samples and buffers were expressed as a function of the modulus of the scattering vector $Q = (4\pi/\lambda)\sin\theta$ with 2θ being the scattering angle and λ the neutron/X-ray wavelength. Buffer intensities were subtracted from the respective sample intensities using PRIMUS⁴⁶. The radii of gyration R_g of all samples were extracted using the Guinier approximation with the same program. The validity of the Guinier approximation, R_g for $Q < 1.3$, was verified and fulfilled in each case. For validation of the crystal structure, CRYSON/CRYSON^{47,48} was employed to fit the back-calculated scattering curves with the experimental SAXS/SANS curves. For structural visualization of the validity, DAMMIN⁴⁹ was used to generate a bead model from the pairwise distribution function derived from the scattering data using GNOM⁵⁰. Ten models have been generated and an average was obtained using DAMAVER⁵¹. This *ab initio* model was superimposed with the crystal structure using SUPCOMB⁵². An *ab initio* bead model has also been generated using MONSA⁴⁹, where SANS data are taken into account. Protein and RNA scattering lengths as well as protein volumes were calculated according to Jacrot⁵³, RNA volumes were calculated according to Voss and Gerstein⁵⁴.

Isothermal titration calorimetry. All titrations were carried out using a MicroCal iTC200 calorimeter (Microcal) at 25 °C. All protein samples were dialysed against NMR buffer without DTT to reduce noise before measurements. The purchased RNA (IBA) was resolubilized in the same buffer to ensure a reasonable baseline. Concentrations for each component in each experiment are listed in Extended Data Table 2. For each titration 40 injections of 1 µl of titrant were made at 120 to 180 s intervals, while stirring at 500 r.p.m. Data were reduced with heat spikes from control experiments, where protein or RNA was titrated into the buffer. Raw data were

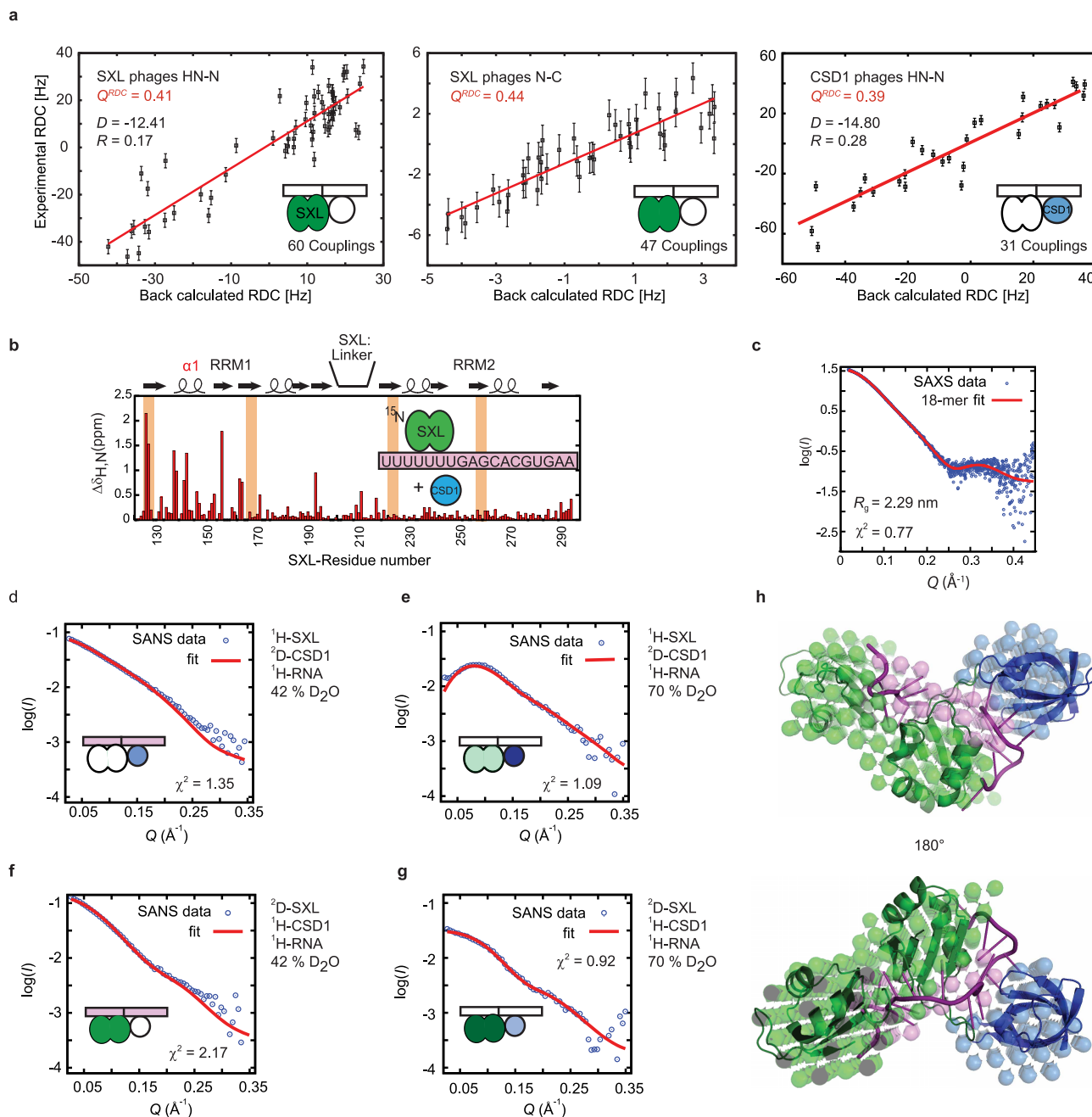
integrated, normalized and the titration curve fitted using Origin isothermal titration calorimetry analysis software provided by Microcal.

30. Hennig, J., Wang, I., Sonntag, M., Gabel, F. & Sattler, M. Combining NMR and small angle X-ray and neutron scattering in the structural analysis of a ternary protein-RNA complex. *J. Biomol. NMR* **56**, 17–30 (2013).
31. Wen, J., Arakawa, T. & Philo, J. S. Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Anal. Biochem.* **240**, 155–166 (1996).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
33. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
35. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321 (2003).
36. Gebauer, F., Corona, D. F., Preiss, T., Becker, P. B. & Hentze, M. W. Translational control of dosage compensation in *Drosophila* by Sex-lethal: cooperative silencing via the 5' and 3' UTRs of *msl-2* mRNA is independent of the poly(A) tail. *EMBO J.* **18**, 6146–6154 (1999).
37. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
38. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wuthrich, K. [13C]-constant-time [15N,1H]-TROSY-HNCA for sequential assignments of large proteins. *J. Biomol. NMR* **14**, 85–88 (1999).
39. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
40. Goddard, T. D. & Kneller, D. G. SPARKY 3. (Univ. California)
41. Cordier, F., Dingley, A. J. & Grzesiek, S. A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and dipolar one-bond J-couplings. *J. Biomol. NMR* **13**, 175–180 (1999).
42. Yang, D. & Kay, L. E. Improved 1HN-detected triple resonance TROSY-based experiments. *J. Biomol. NMR* **13**, 3–10 (1999).
43. Otting, G., Ruckert, M., Levitt, M. H. & Moshref, A. NMR experiments for the sign determination of homonuclear scalar and residual dipolar couplings. *J. Biomol. NMR* **16**, 343–346 (2000).
44. Dosset, P., Hus, J. C., Marion, D. & Blackledge, M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR* **20**, 223–231 (2001).
45. Gosh, R. E. *et al.* A computing guide for small-angle scattering. ILL Technical Report ILL06GH05T. (2006).
46. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277–1282 (2003).
47. Svergun, D. I., Barberato, C. & Koch, M. H. J. CRY SOL - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768–773 (1995).
48. Svergun, D. I. *et al.* Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl Acad. Sci. USA* **95**, 2267–2272 (1998).
49. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
50. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using preceptual criteria. *J. Appl. Cryst.* **25**, 495–503 (1992).
51. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Cryst.* **36**, 860–864 (2003).
52. Kozin, M. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.* **34**, 33–41 (2001).
53. Jacrot, B. Study of biological structures by neutron-scattering from solution. *Rep. Prog. Phys.* **39**, 911–953 (1976).
54. Voss, N. R. & Gerstein, M. Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J. Mol. Biol.* **346**, 477–492 (2005).
55. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**, 6836–6837 (1998).
56. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
57. St Pierre, S. E., Ponting, L., Stefancsik, R., McQuilton, P. & the FlyBase Consortium. FlyBase 102—advanced approaches to interrogating FlyBase. *Nucleic Acids Res.* **42**, D780–D788 (2014).
58. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* <http://dx.doi.org/10.1038/msb.2011.75> (11 November 2011).
59. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
60. Jacques, D. A., Guss, J. M., Svergun, D. I. & Trehwella, J. Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta Crystallogr. D* **68**, 620–626 (2012).



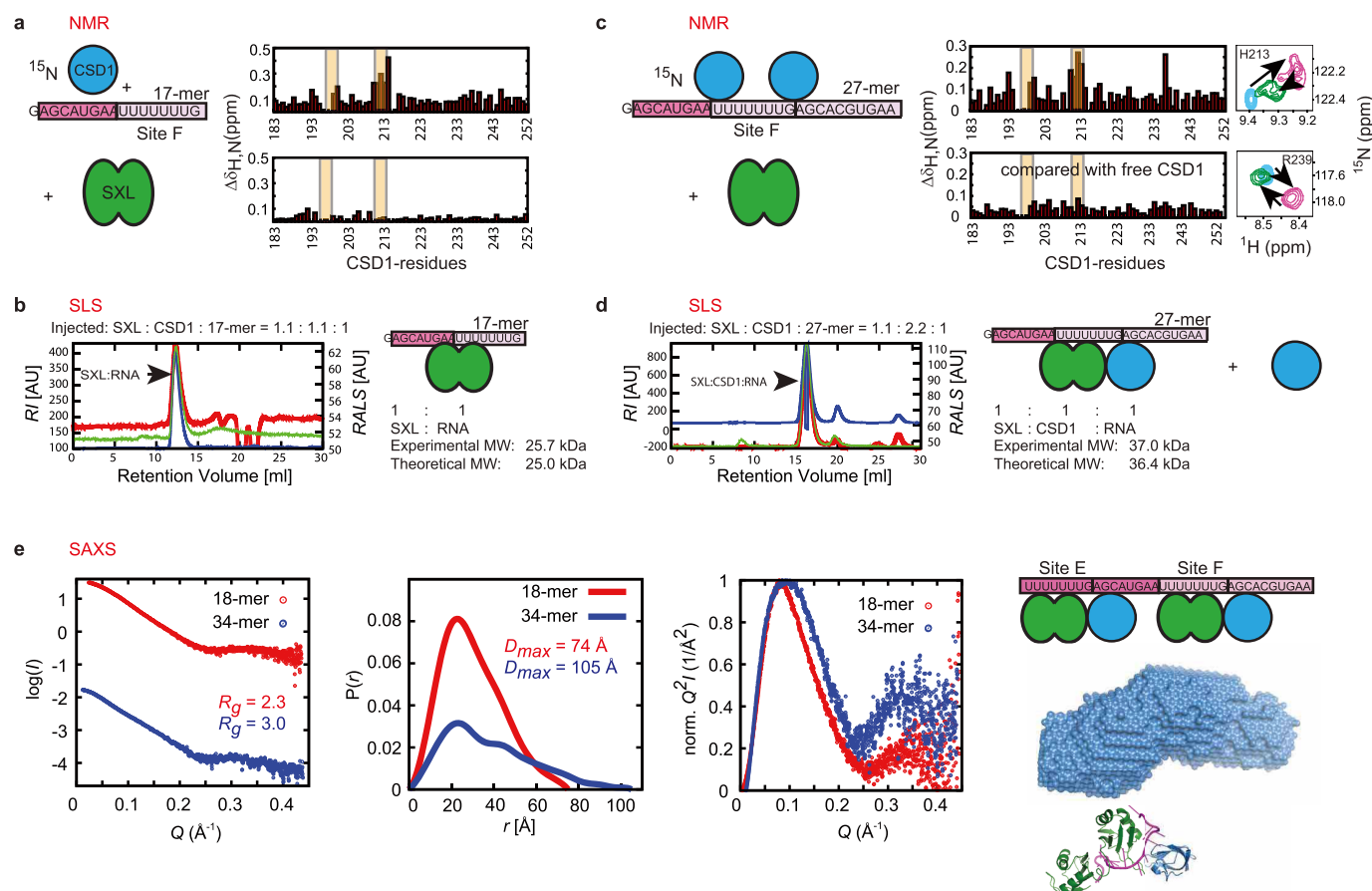
Extended Data Figure 1 | Biochemical and biophysical characterization of Sxl-Unr-*msl2*-RNA complex formation. **a**, Sxl and Unr do not interact in the absence of RNA. The lack of interaction between Sxl dRBD3 and Unr CSD1 in the absence of RNA was tested by an NMR titration, where ^{15}N -labelled Sxl resonances were observed in a ^1H , ^{15}N -HSQC experiment (blue). Upon titration of CSD1 up to a ratio of 1:1 (red), no significant changes of the spectrum could be observed. Some line broadening is observed for Leu 127, Gly 167, Asp 210, Val 238, Arg 244 and Ser 281. However, as these residues are not involved in complex formation, this probably reflects unspecific interactions and/or aggregation. **b**, Mapping of the Unr binding region on *msl2* mRNA using EMSA. Serial mutation of the EF region replacing the native nucleotides by CU repeats showed that the first 32 nucleotides were

relevant (Fig. 1), while the last 14 nucleotides (mutant (Mut) 5 and 6) are dispensable for ternary complex formation. **c**, Size-exclusion chromatography combined with static light scattering for molecular weight (MW) determination of dRBD3-CSD1-*msl2*-RNA complexes using different lengths of RNA. The refractive index (RI, red) and the right-angle light scattering (RALS, green) profiles are shown. AU, arbitrary units. **d**, Structural superposition of Sxl bound to *transformer* pre-mRNA (Sxl, yellow; RNA, orange, PDB: 1B7F¹¹) and bound to *msl2* mRNA in the ternary complex presented in this study (Sxl, green; RNA, magenta). **e**, Comparison of the structure of Unr CSD1 bound to *msl2* within the ternary complex (left: CSD1, blue; RNA, magenta), and the homologous CSD of CspB from *Bacillus subtilis* bound to U5 RNA (right: CSD, orange; RNA, yellow, PDB: 3PF5¹¹).



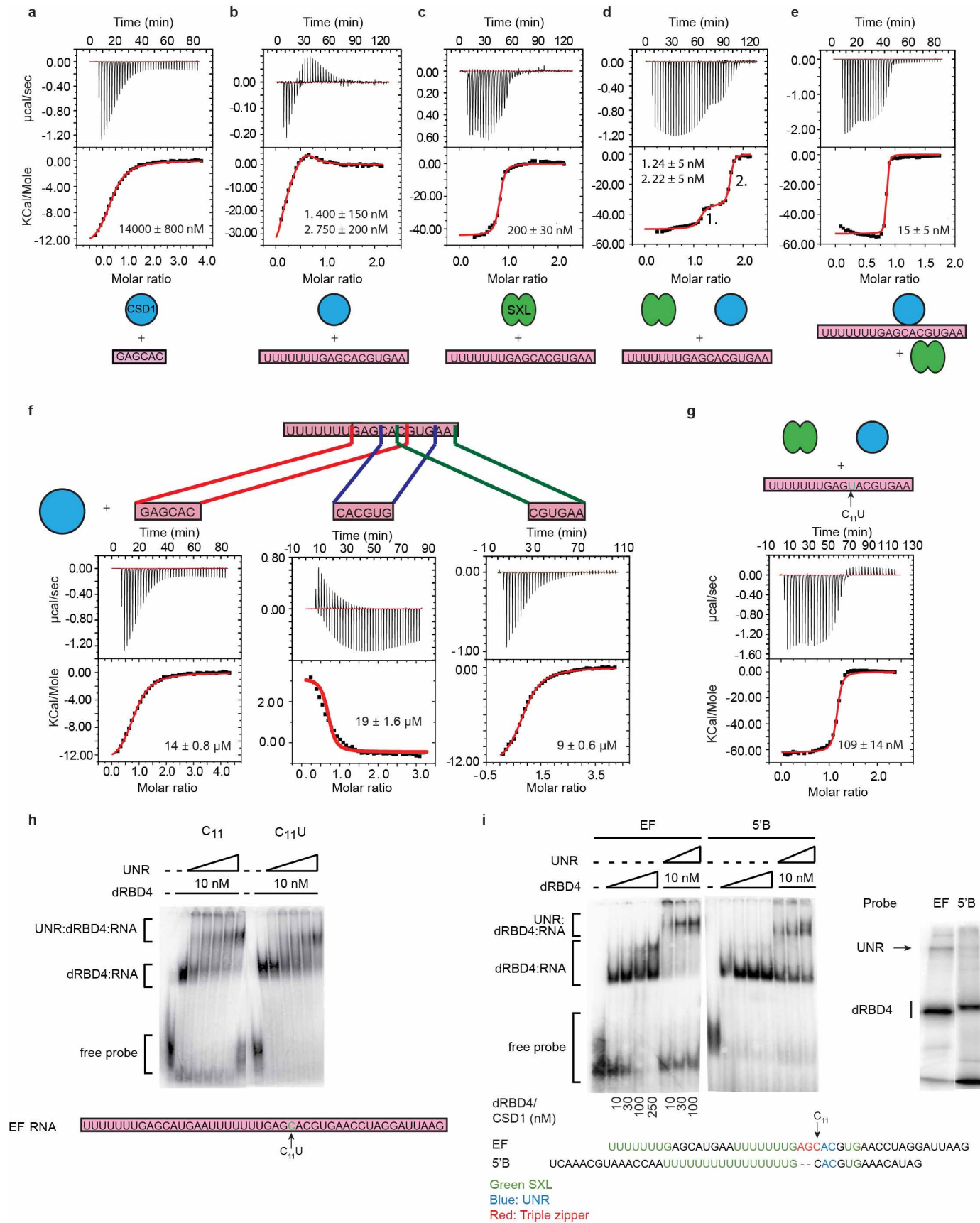
Extended Data Figure 2 | Structure validation in solution. **a**, Residual dipolar couplings (RDCs) validate the domain orientations for the Sxl RRM1 and Unr CSD1 seen in the crystal structure in solution. Correlations between experimental and back-calculated residual dipolar coupling values are shown. The type of dipolar coupling, the observed protein within the complex, and the number of couplings are indicated. All measurements were performed in Pf1 phages. Other alignment media tested induced aggregation of the sample and interacted with the complex as chemical shift perturbations could be observed. The RDC quality factors Q^{RDC} are shown in red and indicate a good agreement between the crystal structure and the domain arrangements in solution. The alignment tensors determined are similar for dRBD3 ($D = -12.4$, $R = 0.17$; where D is the magnitude and R the rhombicity of the alignment tensor) and CSD1 ($D = -14.8$, $R = 0.28$) consistent with a fixed domain orientation in solution. Error bars indicate measurement uncertainties depending on the spectral resolution. **b**, NMR chemical shift perturbation (combined ^1H and ^{15}N chemical shift difference $\Delta\delta_{H,N}$) of the Sxl-RNA complex upon titration with CSD1. Major shifts occur only on residues located in RRM1, consistent with the crystal structure. Further zoom views of triple-zipper and RRM1- $\alpha 1$ -helix residues confirm their involvement in ternary complex formation (Fig. 3a). **c**, Small-angle X-ray scattering (SAXS) analysis of

the ternary complex showing back-calculated scattering densities (red line) and experimental scattering data (blue dots). A χ^2 of 0.77 confirms that the structure of the complex in solution is the same as in the crystal structure. **d**, Small-angle neutron scattering (SANS) of the ternary dRBD3-CSD1-*msl2*-mRNA complex, where CSD1 is perdeuterated and dRBD3/*msl2*-mRNA is protonated, at 42% D_2O buffer concentration. In this composition, scattering of dRBD3 matches the scattering density of the buffer and therefore does not contribute to the signal (indicated with a white dRBD3 component in the schematic complex used throughout this study). The back-calculated scattering curve (red line) is fitted against the experimental data (blue dots), and the resulting χ^2 is shown as inset. **e**, as **d** but using 70% D_2O in the buffer, where RNA matches the buffer contrast, and where the perdeuterated component (CSD1) has a positive contrast (indicated with a dark blue colour), and dRBD3 a negative contrast (light green). **f**, as **d** but here dRBD3 is perdeuterated and CSD1 is protonated. Thus, CSD1 matches the contrast of the buffer and does not contribute to the signal, whereas dRBD3 exhibits a positive contrast. **g**, as **e** but inversely labelled, thus, CSD1 having a negative contrast to the buffer and dRBD3 a positive. These data enabled us to localize each component within the overall *ab initio* bead models as shown in **h** obtained with the program MONSA⁵⁰ (two views are shown; green, Sxl; blue, CSD1; magenta, RNA; see Methods for details).



Extended Data Figure 3 | NMR and small-angle X-ray scattering analysis of dRBD3–CSD1 bound to different *msf2* mRNA oligomers. **a**, NMR chemical shift perturbations (combined ^1H and ^{15}N chemical shift difference $\Delta\delta_{\text{H,N}}$) of amide signals of CSD1 upon binding to 17-mer RNA (Site F with the 5' flanking region) (top right histogram; conserved, predicted RNPs are indicated by yellow bars). Very similar chemical shift perturbations as for the 18-mer (Site F with 3' flanking region) are observed (Fig. 3, main text). However, upon further titration with dRBD3 no additional shifts can be seen (bottom histogram), suggesting that there are no contacts between CSD1 and Sxl, in contrast to the 18-mer. **b**, Static light scattering (SLS) confirms these observations and further shows that a stable complex can be only formed between Sxl and the 17-mer RNA (RALS: right-angle light scattering; RI: refractive index). **c**, as in **a** but instead of a 17-mer, a 27-mer RNA consisting of site F and both flanking regions was used. This RNA can accommodate two CSD1 moieties as evident from NMR chemical shift perturbations (top histogram). However, upon Sxl titration, the NMR signals move back to the position of the free CSD1 (bottom histogram), as exemplified by the chemical shift changes observed for the His 213 and Arg 239 triple-zipper residues (zoomed spectra on the right). This shows that at least one CSD1 is pushed out of the complex with the 27-mer. **d**, These results were confirmed by static light scattering, where only a stable 1:1:1 complex (Sxl–CSD1–RNA) could be formed. The experimentally determined molecular weight is consistent with a 1:1:1 complex. **e**, Small-angle X-ray scattering (SAXS) data of the 18-mer (red) compared with the 34-mer (blue) suggest that two Sxl–CSD1 complexes are independently assembled on the 34-mer like 'beads on a string', that is, with no contacts between the two subcomplexes bound to site E and F, respectively. The radius of gyration is

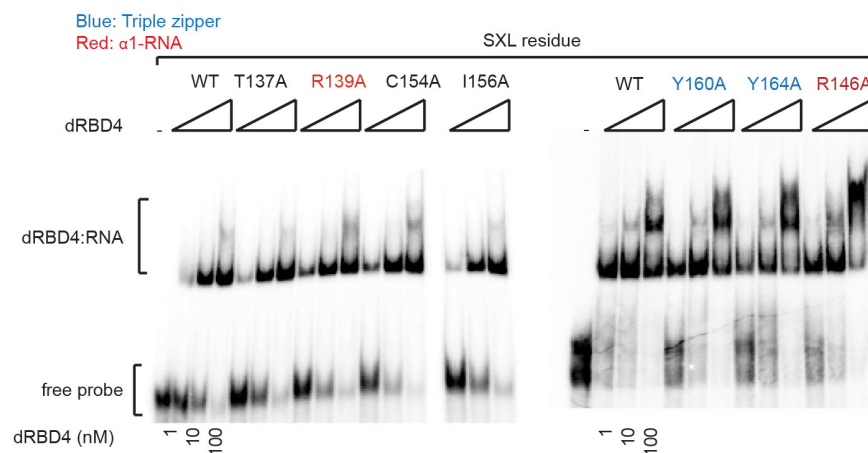
much larger and D_{max} (the maximum diameter of the particle, derived from the distance distribution function $P(r)$) increases more than would be expected for a compact and rigid 34-mer complex. Although Kratky plots are consistent with a structured 34-mer complex, the curve does not reach down to the baseline after the peak as for the 18-mer complex, indicating a higher degree of flexibility. I , Q , r and R_g refer to the scattering intensity, the scattering vector, interatomic distance and radius of gyration, respectively. A shape analysis, using *ab initio* modelling (see Methods for details), demonstrates that the average of 20 models adopts an elongated shape (right) to accommodate two Sxl–CSD1 complexes bound to the 34-mer RNA. The data presented in this figure, the inability to produce crystals in extensive crystallization trials, and the difference in RNA binding affinity for sites E and F (see main text, Fig. 1), as well as the non-conserved distance between the E and F sites in different *Drosophila* species (Extended Data Fig. 7) suggest the lack of contacts between both single complexes (site E and site F complex), at least in the absence of other co-factors. Notably, the weaker affinity of the E site is probably due to the missing CG dinucleotide in the corresponding 3' flanking regions (AGCAUGAA flanking E site vs AGCAGUGAA flanking F site). The C11 nucleotide in the 18-mer (flanking the F site) is specifically recognized in the 18-mer complex (Fig. 2). In the E site this contact would be formed with a U instead of a C, thereby missing the hydrogen bond between the NH_2 of the base with the backbone carbonyl oxygen of Sxl Lys 161. As a result, the subsequent AA dinucleotides might not be able to form the non-canonical contacts with the $\alpha 1$ helix of Sxl RRM1 and thus maybe induce the flexibility observed in SAXS.



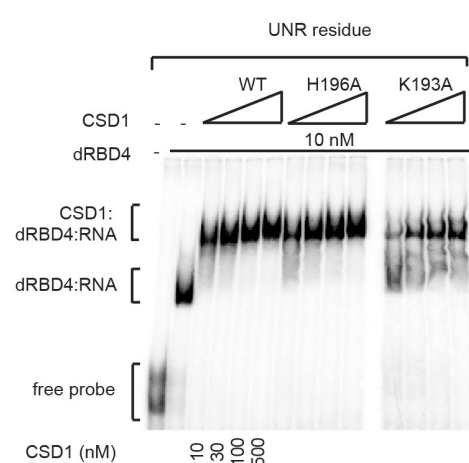
Extended Data Figure 4 | Isothermal titration calorimetry analysis of the dRBD3–CSD1–*msl2*-mRNA interactions. **a**, Isothermal titration calorimetry data of the cognate RNA (a hexamer, including the bases bound to CSD1 in the ternary complex) titrated to CSD1 can be fitted to an affinity corresponding to a dissociation constant K_d of 14 μ M. **b**, The 18-mer can accommodate two CSD1 moieties which bind with similar affinity. The second binding site is within the exothermic part of the isothermal titration calorimetry binding curve, which arises due to the simultaneous melting of the self-complementary part of the RNA (CACGUG). This can also be observed, when titrating a hexamer oligonucleotide consisting of this sequence alone into CSD1 (**f**). The higher affinity compared to the hexamer is due to the increased availability of binding surface. **c**, Isothermal titration calorimetry data of titrating dRBD3 into an 18-mer. Fitting the data, assuming one binding site yielded dissociation constant K_d of 200 nM, similar to other observations¹⁶. **d**, Titration of 18-mer into a sample containing equimolar amounts of dRBD3 and CSD1 exhibits two binding sites with similar affinities (K_d = 15–20 nM). **e**, Titration of dRBD3 into a sample containing 18-mer–CSD1 complex resulted in one binding site with an affinity (a K_d) of 15 nM. Thus, the affinity of CSD1 in the context of the complex increased 1,000-fold compared to its affinity to RNA alone. **f**, Specificity of *msl2* mRNA recognition by Sxl and CSD1. Isothermal titration calorimetry data shows that CSD1 has similar weak affinities to different hexamer subsets of the 18-mer in the range of K_d = 10–20 μ M (the exothermic signals in the middle are due to simultaneous melting of the self-complementary

oligonucleotide CACGUG), although CSD1 specifically recognizes C11 in the context of the 18-mer complex in presence of Sxl. Initial specificity seems to be governed by Sxl which binds to 18-mer RNA with higher affinity and recruits CSD1 to its position with extremely increased affinity (1,000-fold), where it, together with Sxl, specifically recognizes C11 as described in the main text. The RNA sequence recognized by UNR CSD1 in the ternary complex (G10–C11–A12–C13–G14) resembles the proposed consensus sequence (A/GAACG/A) for the human homologue²¹. However, the crucial C11 is not present in the consensus. A reason for this could be that a Sxl-like binding partner was not present in the SELEX experiments; alternatively, human UNR could show a different binding specificity. **g**, Isothermal titration calorimetry of ternary complex formation with C11U mutant. Affinity decreases by fivefold if the C11 is substituted by a uridine, which is in agreement with EMSA data (**h**). The two transitions observable for wild-type 18-mer are not visible here. This is probably due to similar affinities for complex formation and Sxl's affinity to the 18-mer alone. Numbers indicated are dissociation constants (K_d). Errors indicate the fitting error. For each experimental setup at least duplicates have been acquired. **i**, Complex formation at site B in the 5' UTR of *msl2*. Binding of dRBD4 and Unr to the E/F site and 5'B probes was assessed by EMSA (left) and crosslink-immunoprecipitation of Unr from embryo extracts (right). Probe sequences are indicated at the bottom. Probe 5'B does not support efficient complex formation *in vitro* or in extracts.

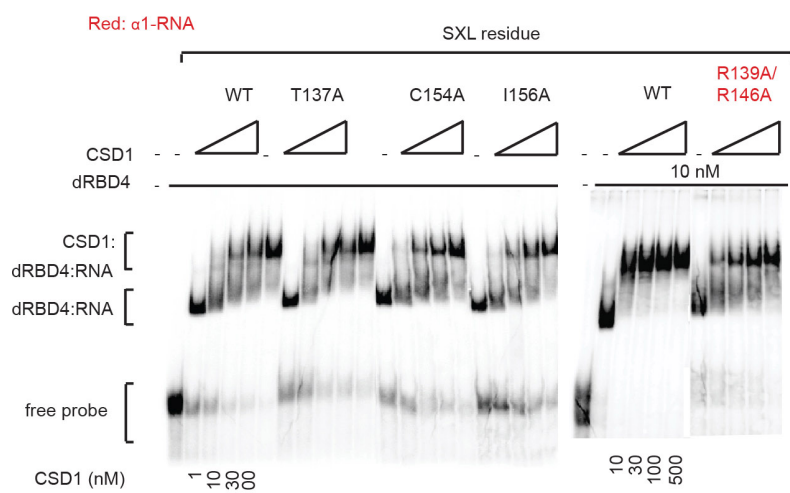
a Binding of SXL dRBD4 mutants to RNA (binary complex)



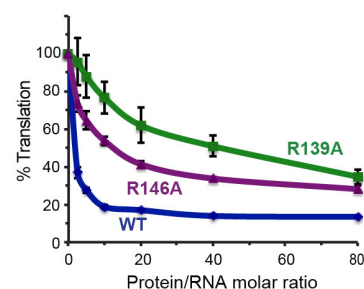
b UNR CSD1 mutations in ternary complex formation



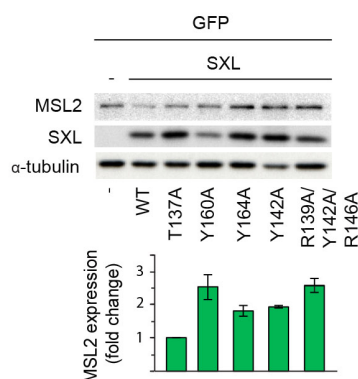
c SXL dRBD4 mutations in ternary complex formation



d Translational repression

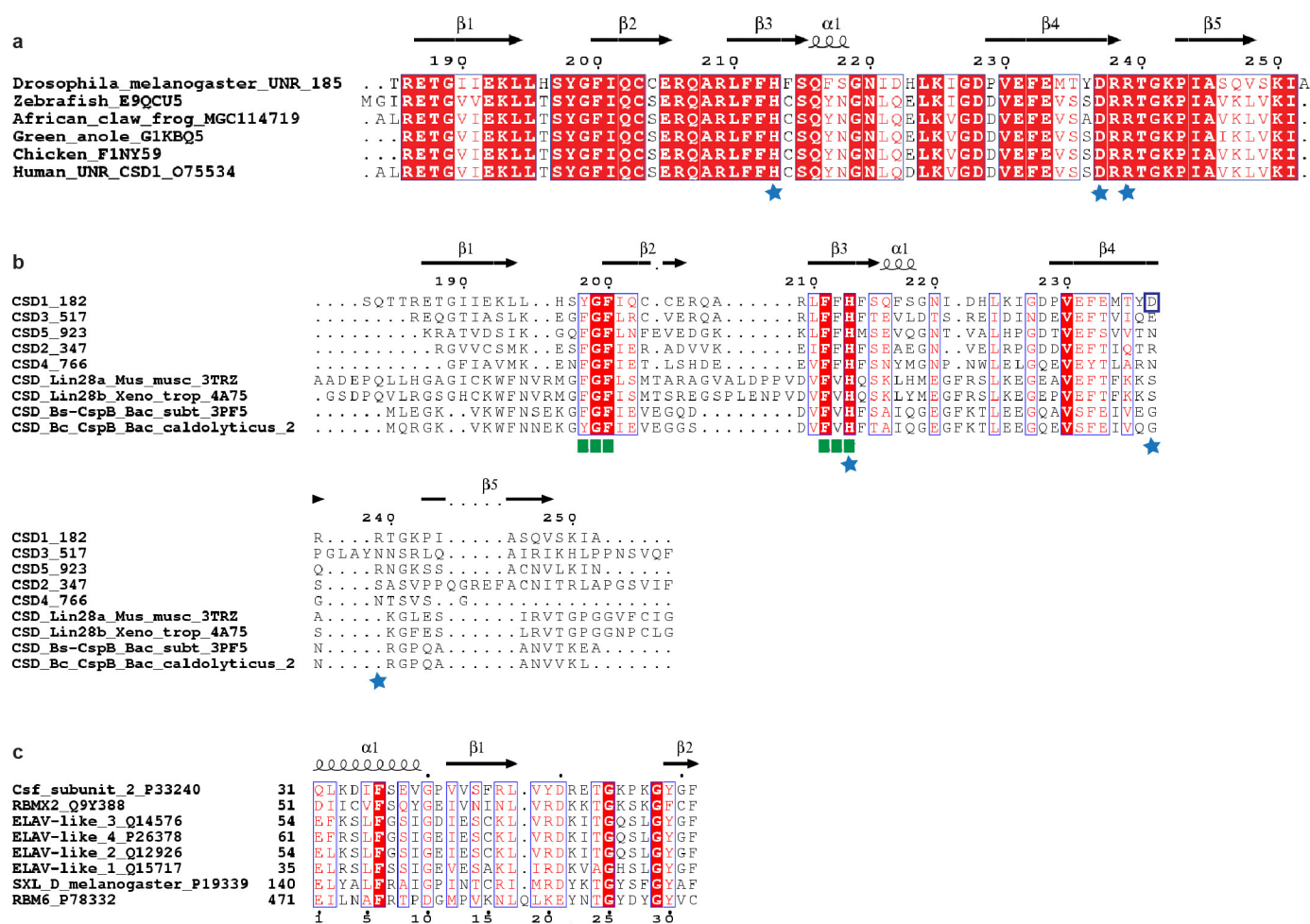


e MSL2 expression in S2 cells transfected with SXL



Extended Data Figure 5 | Electrophoretic mobility shift assays of dRBD4 and CSD1 variants. **a**, Assessment of the RNA binding ability of dRBD4 mutants. All mutants exhibited the similar binding patterns as the wild type (WT), indicating that all mutants fully retained their RNA binding capacity. **b**, Assessment of the ability of dRBD4 mutants to form a ternary complex with RNA and CSD1. **c**, Assessment of the ability of CSD1 mutants to form a ternary complex with RNA and dRBD4. **d**, Translational repression assays in embryo extracts (containing endogenous Unr) and increasing amounts of dRBD4

derivatives. A luciferase reporter containing the *msl2* EF region in the 3' UTR was used (BmutLEF). Error bars represent the standard deviation of three experiments. **e**, Analysis of Msl2 expression in Schneider 2 cells transfected with Sxl derivatives. Msl2 and Sxl expression were monitored by western blot (upper panel). Msl2 was quantified, normalized for Sxl expression, and reported relative to the T137A negative control. Error bars represent the standard deviation of two experiments.



Extended Data Figure 6 | Multiple sequence alignment of cold shock domains and human Sxl orthologues. **a**, Sequence alignment of *Drosophila* Unr CSD1 with CSD1 of Unr of other species. Completely conserved residues are framed and have a red background. Residues which are exchanged by a similar residue are framed and have a red font. The secondary structure of CSD1 in the crystal structure presented here is indicated above the alignment. Unr CSD1 is highly conserved even between different phyla (arthropoda vs chordata) and within the different classes of chordata (actinopterygii, amphibia, reptilia, aves and mammalia). Residues involved in complex formation (His 213, Asp 237 and Arg 239, indicated by a blue star) are fully conserved, suggesting that CSD1 of human Unr might also form a similar complex with human counterparts of Sxl in other pathways. **b**, Alignment of CSD1, CSD2, CSD3, CSD4 and CSD5 of *Drosophila* Unr. The starting residue number of each domain within full length Unr is in front of each sequence. Same colour

code as in **a**. The highly conserved RNP1 and fully conserved RNP2 are indicated with a green square, indicating that all five cold shock domains are likely to be able to bind RNA. Asp 237 can only be found in CSD1, and Arg 239 in CSD1 and CSD5 (blue stars). **c**, Alignment of RRM1 of cleavage stimulation factor subunit 2 (CstF), RNA binding protein X-linked 2 (RBMX2), four ELAV-like protein homologues (ELAV-like 1–4, ELAV-like 1 = HuR), Sxl and RNA binding motif 6 (RBM6). Only the sequence region containing residues involved in ternary complex formation (α 1 helix, triple-zipper residues) was aligned and was used in an extensive similarity search of human proteins. RBM6 shows the best conservation of triple-zipper (blue star) and α 1-helix (red triangle) residues. Arg 139, the mutation of which has the weakest effect on complex formation is not conserved in any of the listed proteins. ESPript was used in this figure with the default colour code (fully conserved, white font, red background; weakly conserved, red font)⁵⁶.

1 10 20 30 40 50 60

D_sechellia AGAG TAAAGCA TAAGTCAAAA TAAA..GATTA TTTTCGTTTA.....ACGAAAGCGATTATATAG.T.T
D_melanogaster AGAG TAAAGCA TAGTCAAAA TAAA..GATTA TTTTCGTTTA.....ACGAAAGCGATTATATAG.A..
D_yakuba AGAG CAAAGCA ATAGTCAAAA TAAA..GATTA TTTTCGTTTA.....ACGAAAGCGATTATATAG.AAT
D_erecta AGAG CAAAGCA ATAGTCAATA TAAA..TATTA TTTTCGTTTA.....ACGAAAGCGATTATATAG.AAT
D_pseudoobscura TTTT CATAGCT TAG.....GTA.....
D_persimilis TTTT CATAGCT TAG.....GTA.....
D_willistoni TGAA CAAAGCAT GTTCTATTAA CAAAAATATGATGTCATATA.....TATATCAACTAAAA..
D_mojavensis TTTT CATAGCTTGT...AAGGC...AATTTTTCATAGATGTAAGAAATTTTGTATAGCTATAAGATAATTT
D_virilis TTTG CATAGCTTAAAG.GATACCTT...TTTTCATATA.....ATTATAGCAATTTTTT...
D_grimshawi ..GACAAAGAAATAT.AAATGTCGAAAAATTTGATTTTCGTATA..ATAGAAA...TTAAATCA..ATTT.....TT

70 80 90 100 110

D_sechellia TACTTAAATGCTTTGCGCAAAATGC.....GATGC...TGTG.AGCTAAA..TAACTTTT.TTT
D_melanogasterTTGCTTTGCGCAAAATGT.....GATGC...TGTG.AGCTAAA..TAACTTTT.TTT
D_yakuba TACTTAAATGCTTTGCGCAAAATGC.....GATGC...TGTG.AGCCGGA..AAACCTTTT.TTT
D_erecta TACTTAAATGCTTTGCGCAAAATGC.....GATGC...TGTG.AGCCGGA..AAACCTTTT.TTT
D_pseudoobscuraTTTTCATAGCT.TAAGTGCAATTTTTCATATTTAAGTACA...TTT.TTT
D_persimilisTTTTCATAGCT.TAAGTGCAATTTTTCATATTTAAGTACA...TTT.TTT
D_willistoniGTGTAT.AAACTAATTTTTCATATCTCTAAGACAATTTTTCATATCTAAGACAA...TTT.TTT
D_mojavensis TTTTCATAAGCTTAAGCAATTTT.TTTA...ATATTTTAAAGCGCAAGTTTGTTCATTAGTTTAAAGCGGATATTTT.TTT
D_virilis .TC..ATATGTTTAAAGCAATTTT.TTTTTCATATTTTAAAGTACA...TGTTCATACTATTAGGTCAC...TTT.TTT
D_grimshawi TGTCAATACGCCTAAGCATTTT.TTTTTCATATTTTAAAGAGA.....AGTTT.TTT

E site

120 130 140 150 160

D_sechellia AATAAGCTAA.G.ACATTTTTTTGATAGGCTAAGAA..CTTTTTTTGAGCATGAA.....
D_melanogaster AATAAGGCTAA.G.ACATTTTTTTGATAGGCTAAGAA..CTTTTTTTGAGCATGAA.....
D_yakuba TGTAAATCTAA.G.AAATTTTTTTGATAGGCTAAGAA..CTTTTTTTGAGCATGAA.....
D_erecta CCTAAGCTAA.G.AAATTTTTTTGATAGGCTAAGAA..CTTTTTTTGAGCATGAA.....
D_pseudoobscura CATAGCTTAAAGTACAAATTTTTTTCATAGATTAAGAA..CTTTTTTTGAGCATGAA.....
D_persimilis CATAGCTTAAAGTACAAATTTTTTTCATAGATTAAGAA..CTTTTTTTGAGCATGAA.....
D_willistoni TTAACATTAAGTACAAATTTTTTTTCATAGGCTAAGAA..CTTTTTTTGAGCATGAA.....
D_mojavensis CATATCTTTTAAAGCAAA.....CTTTTTTTGAGCATGAA.....
D_virilis CATAGTTTAAAGCAAACTTTTTTTTATAGGCTAAGAA..CTTTTTTTGAGCATGAA.....
D_grimshawi CATATCTTTAAGTCACTTTTTTTTTCATAGGCTAAGAA..CTTTTTTTGAGCATGAA.....

F site

170 180 190 200

D_sechelliaTTTTTTTGGACACGTGAACTTAGGA.....TTAAGAA.....CAAC
D_melanogasterTTTTTTTGGACACGTGAACTTAGGA.....TTAAGAA.....CAAC
D_yakubaTTTTTTTGGACACGTGAACTTAGGA.....TTAAGAA.....CAAC
D_erectaTTTTTTTGGACACGTGAACTTAGGA.....TTAAGAA.....C
D_pseudoobscuraTTTTTTTGGACACGTGAACTAATATTAATA.....
D_persimilisTTTTTTTGGACACGTGAACTAATATTAATA.....
D_willistoniTTTTTTTGGACACGTGAACTAATATTAATA.....
D_mojavensis ACGAAACCCCTTTTTTTTGGACACGTGAACTTTTTTTTTTATT.....ATTATTTAAGATTTGTGTATGAATT
D_virilis ATGA...ACCTTTTTTTTGGACACGTGAACTCAACATTTTTTTTTT...TTCTTAATTTAAGATTGTGTATCTCG
D_grimshawi ATGAAACC...TTTTTTTGGACACGTGAACTCAATTTTTTTTTTCA.....TAGTTTAAGATATTTTATCTCG

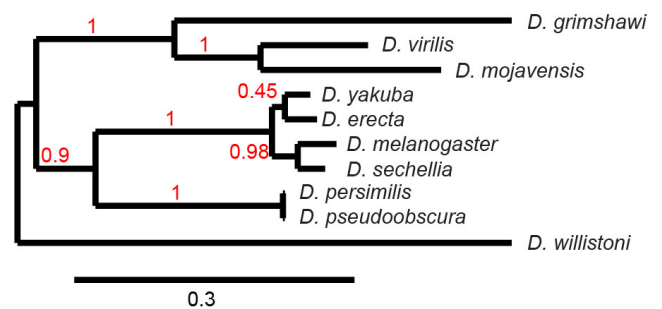
c¹¹ ↑

210 220 230 240 250

D_sechellia AATTCAGTAAGCCATATATACCATAT...TATATATGCGAAATCAT.....ATAACTATTAA.....
D_melanogaster AATTCAGTAAGCCATATATACCATAT...TATATATGCGAAATCAT.....ATAACTATTAA.....
D_yakuba AATTCAGTAAGCCATATATACCATAT...TATATATGCGAAATCAT.....ATAACTATTAA.....
D_erecta AATTCAGTAAGCCATATATACCATAT...TATATATGCGAAATCAT.....ATAACTATTAA.....
D_pseudoobscuraAGCCATATAA.....TATGTAAATGCTAACATATAAATT.CAAA.....ATATGCGAAACCA
D_persimilisAGCCATATAA.....TATGTAAATGCTAACATATAAATT.CAAA.....ATATGCGAAACCA
D_willistoni GCAAAATATAATCCACATATAAATTAACCTATGTAAATAAATGGAATAATTT.TGATA..CATTTATTCGAT...
D_mojavensis GAATTTTA.AAATCCTAATTG..CA..AAAAATGTAATGCAATTTGAACAATAA...AAATGAACCTA.AAATCTAAACA
D_virilis GAATTTGAAGCCAGTTGCAATATAA..TATTTGTTTGCATAATTTAAATAAAG...AGATTTTCTAACGCAAT...
D_grimshawi CAGTTATGAAGTAATTTG...CATATAAGATAATTTTGCATAATTTAATACTATACGATAATGAACCTTAACCTTAAGCA

260

D_sechelliaAAATTTTAA
D_melanogasterAAATTTTAA
D_yakubaAAATTTTAA
D_erectaAAATTTTAA
D_pseudoobscura ATATA.....TGATATGTA
D_persimilis ATATA.....TGATATGTA
D_willistoniGTAATGTC
D_mojavensis ATTTGT.....TAAATCCGTTAGGTGTGCA
D_virilisTCAGATTGAATTTAA
D_grimshawi AAAAAAAAAAAACAAACAAACCAATGAAATCTGAAT



Extended Data Figure 7 | Conservation of the 3' UTR of *msl2* mRNA between different *Drosophila* species. Alignment of a region of the 3' UTR of *msl2* mRNA of different *Drosophila* species indicates that the E and F sites, where the ternary complex is formed, are highly conserved, especially C11 (indicated with an arrow), which is fully conserved. The flanking regions show much less conservation and regions towards the 5' and 3' ends of the sequences aligned here are even less conserved (not shown). A phylogenetic tree is shown, illustrating the degree of diversion of the species used in the alignment. The

variable length of the region between sites E and F is further indirect evidence for the lack of contact between complexes formed on each site (Extended Data Fig. 3). The sequences were retrieved from <http://flybase.org>⁵⁷. The multiple sequence alignment was performed using Clustal Omega⁵⁸, and the phylogenetic tree was generated by Phylogeny.fr⁵⁹. The alignment figure was generated using ESPrnt with the default colour code (fully conserved, white font, red background; weakly conserved, red font)⁵⁶.

Extended Data Table 1 | Crystallographic data collection and refinement statistics

	SXL:CSD1:18-mer
Data collection	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	94.4, 111.0, 139.2
Resolution (Å)	50-2.8 (2.91-2.84) *
<i>R</i> _{meas}	5.7 (87.3)
<i>I</i> /σ <i>I</i>	21.62 (2.88)
Completeness (%)	99.8 (99.4)
Redundancy	7.4 (7.1)
Refinement	
Resolution (Å)	20-2.8 (2.91-2.84)
No. reflections	24847
<i>R</i> _{work} / <i>R</i> _{free}	20.7/23.9 (40.6/45.0)
No. atoms	
Protein	7704
RNA	1076
Water	0
B-factors	
Protein	101.4
RNA	94.39
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.05
R.m.s deviations between two complexes in asymmetric unit (Å)	
All atoms	1.41 ^s
Cα	0.77

The highest resolution shell is shown in parentheses. Both complexes in the asymmetric unit are near identical, except Sxl Tyr 164, which adopts in one complex the native contact with RNA as in the crystal structure of Sxl bound to *transformer* pre-mRNA. In the other complex however, Tyr 164 flips towards CSD1 Arg 239 and forms part of the triple zipper (Fig. 2b). This might be due to the non-native N-terminal GAMA residues (remaining from the TEV cleavage site) of CSD1, which form crystal contacts between Tyr 164 and RNA. However, the dual conformation of Tyr 164 and its participation in the triple zipper is confirmed by its strong NMR chemical shift perturbation upon CSD1 binding and the fact that the Sxl mutant Y164A retains its complete RNA binding capacity but shows weakened complex formation with CSD1.

Extended Data Table 2 | Small-angle scattering statistics according to Jacques *et al.*⁶⁰

	18-mer complex	34-mer complex	18-mer complex
Data collection			
Instrument	BioSAXS BM29 ESRF	Anton Paar GmbH, SAXSess mc ²	SANS D22 ILL
Beam geometry	10 mm slit	Slit collimation	Pin hole
Wavelength (Å)	1.008	1.5	6
Q range (Å ⁻¹)	0.004-0.45	0.019-0.44	0.027-0.35
Exposure time (s)	2	5400	60-3600
Concentration range (mg ml ⁻¹)	0.5-3.6	0.75-3.0	3.6-10.5
Temperature (K)	298	298	298
Structural parameters			
$I(0)$ (cm ⁻¹) from $P(r)$	39.3 ± 0.040	0.018 ± 0.000	0.142, 0.035, 0.084, 0.008 (± 0.001) *
R_g (Å) from $P(r)$	23.0 ± 0.003	30.0 ± 0.007	22.8 ± 0.2, 16.9 ± 0.9, 20.9 ± 0.4, N.D. *
$I(0)$ (cm ⁻¹) from Guinier	38.9 ± 0.34	0.017 ± 0.001	0.142, 0.036, 0.084, 0.012 (± 0.001) *
R_g (Å) from Guinier	22.8 ± 0.30	29.6 ± 0.35	22.1 ± 0.3, 18.2 ± 1.2, 20.0 ± 0.5, 22.3 ± 1.2 *
D_{max} (Å)	74	105	78, 78, 78, 78 *
Porod volume estimate (Å ³)	55300	125360	N.D.
Molecular mass determination			
Molecular mass from $I(0)$ (kDa)	34.8 †	80.0 †	38.0 ± 4 †
Theoretical mass from sequence (kDa)	33.4	66.2	33.4
Software employed			
Primary data reduction	BsxCuBE	SAXSQuant (v3.9)	ILL inhouse software
Data processing	PRIMUS	PRIMUS	PRIMUS
<i>Ab initio</i> analysis	DAMMIN	DAMMIF §	MONSA
Validation and averaging	DAMAVR	DAMAVR	N.D.
Rigid-body modeling	-	-	-
Computation of model intensities	CRY SOL	CRY SOL	CRYSON

For SANS data, each $I(0)$ and R_g value are listed in the following order: ²H-Sxl-¹H-CSD1 at 42% D₂O, ²H-Sxl-¹H-CSD1 at 70% D₂O, ¹H-Sxl-²H-CSD1 at 42% D₂O, and ¹H-Sxl-²H-CSD1 at 42% D₂O. For SAXS measurements the $I(0)$ has been calibrated for molecular weight determinations against bovine serum albumin. For SANS data, the molecular weight of the 18-mer complex has been calculated only for a fully protonated sample at 0% D₂O. The *ab initio* DAMMIF shape analysis of the 34-mer complex has been performed by calculating 20 models having a spatial discrepancy and averaged normalized standard deviation of 0.751 ± 0.034 . One model has been discarded.

Extended Data Table 3 | Isothermal titration calorimetry data of the dRBD4-CSD1-*msl2*-mRNA complex

	K_d (M)	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/K · mol)	N	Conc. CSD1 (mM)	Conc. SXL (mM)	Conc. RNA (mM)
6-mer (GAGCAC) → CSD1	$(1.4 \pm 0.3) \cdot 10^{-5}$	-6.7	-13.8	-23.9	1	0.06	-	2.2
18-mer → CSD1						0.06	-	0.6
<i>Site 1</i>	$(3.8 \pm 1.5) \cdot 10^{-7}$	-8.9	-154.4	-488.0	0.5			
<i>Site 2</i>	$(7.5 \pm 2.0) \cdot 10^{-7}$	-8.3	39.4	160.0	0.5			
SXL → 18-mer	$(1.9 \pm 0.3) \cdot 10^{-7}$	-9.1	-44.0	-117.0	1	-	0.6	0.06
18-mer → SXL/CSD1						0.06	0.06	0.6
<i>Site 1</i>	$\sim(2.4 \pm 1.0) \cdot 10^{-8}$	-8.5	233.0	815.0	1			
<i>Site 2</i>	$\sim(2.2 \pm 0.5) \cdot 10^{-8}$	-10.2	-292.0	-945.0	1			
SXL → 18-mer/CSD1	$\sim(1.5 \pm 1.0) \cdot 10^{-8}$	-10.6	-52.9	-142.0	1	0.06	0.49	0.06
18-mer (C¹¹A) → SXL/CSD1	$\sim(6.7 \pm 0.7) \cdot 10^{-8}$	-9.7	-56.5	-157.0	1	0.053	0.053	0.6
18-mer (C¹¹U) → SXL/CSD1	$\sim(1.1 \pm 0.14) \cdot 10^{-7}$	-9.5	-61.9	-176.0	1	0.053	0.053	0.6
18-mer → SXL								
CSD1 → 18-mer				Aggregation				
SXL/CSD1 → 18-mer				Aggregation				
CSD1/18mer → SXL				Aggregation				

TECHNOLOGY FEATURE

A DEEP LOOK AT SYNAPTIC DYNAMICS

The processes behind neuronal communication have not yet been resolved in detail, but dyes, microscopy and protein analysis are beginning to fill in the gaps.



BRAND X PICTURES/GETTY

Synapses are crucial to the communication between neurons, but the events that happen there have been difficult to capture.

BY VIVIEN MARX

Kiss-and-run sounds like a schoolyard prank, but it is also the informal name for one of four vigorously debated hypotheses about what happens in neurons in the brain before and after they transmit signals to one another at the cell-to-cell junctions called synapses.

There are myriad ways in which this delicate messaging can be upset. Drugs such as cocaine

or methamphetamine increase the release of the neurotransmitter dopamine, for example. Disorders such as Parkinson's disease involve damage or destruction of the neurons that release dopamine. And depression is linked to altered levels of neurotransmitters such as dopamine and serotonin.

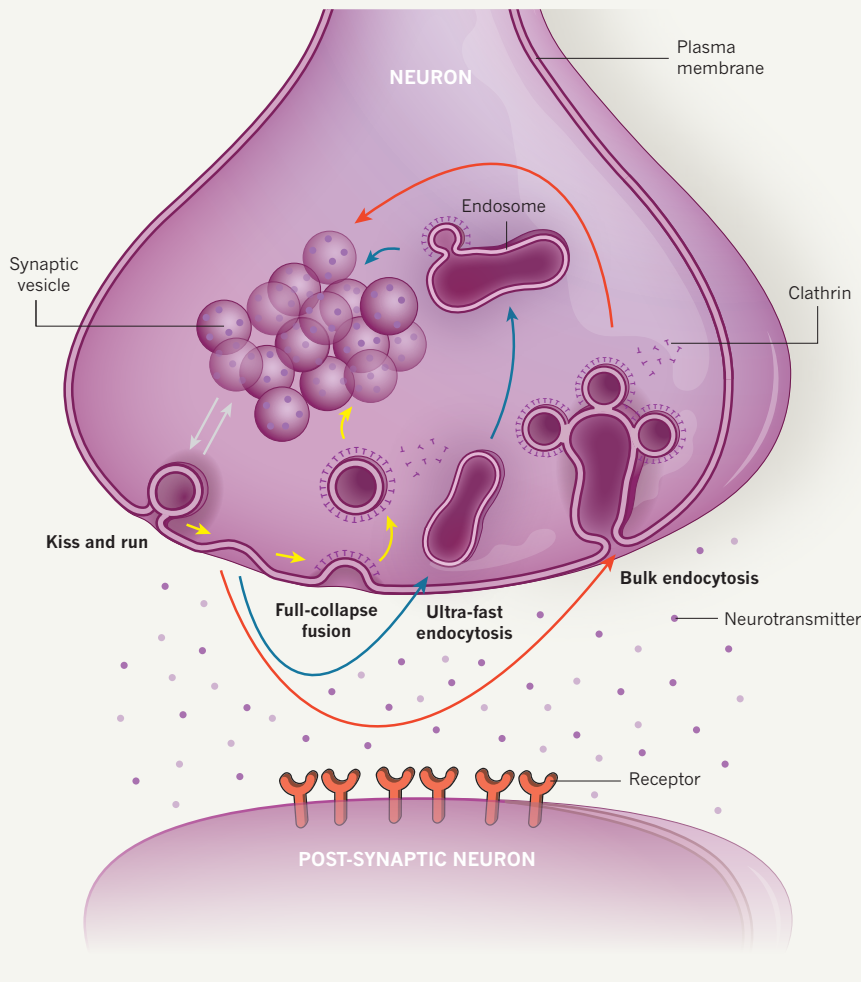
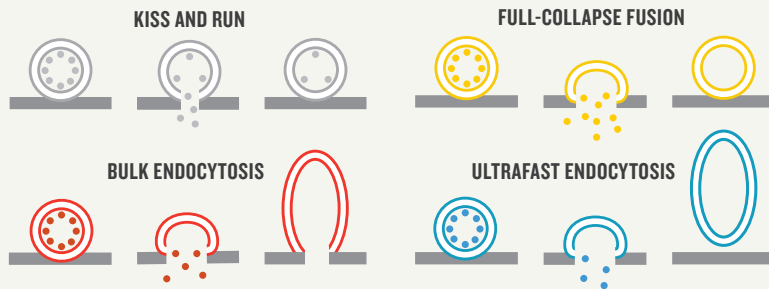
A better understanding of synaptic events is crucial to fighting a wide range of disorders, from drug addiction to mental illnesses — to say nothing of forging a better understanding

of the brain. Scientists particularly want to see how the neurotransmitters travel within the synapse and how they are launched on their journey to the neuron on the other side.

Investigators agree on the basics. A neuron packages its neurotransmitters inside vesicles, which are bubble-like entities around 50 nanometres in diameter. Vesicles are not unique to neurons — they shuttle molecules around in many of the body's cells. But a single neuron can have hundreds of thousands of vesicles, ►

FIRING FOUR

Neurotransmitters are packaged in synaptic vesicles. When called into action, the vesicles move and release their cargo. How that transpires is vigorously debated and studied. There are four hypotheses.



Kavalali of the University of Texas Southwestern Medical Center in Dallas.

There are four major hypotheses. Kiss-and-run, in which the vesicle empties its cargo through a pore in the membrane, then retreats back into the neuron; full-collapse fusion, in which the vesicle melds with the membrane and a new one emerges with the help of a protein called clathrin; bulk endocytosis, in which numerous vesicles fuse with the membrane, which then forms a 'bleb', or bulge, that pinches off to form new vesicles; and ultrafast endocytosis, which is essentially a sped-up version of bulk endocytosis in which pinching off happens in milliseconds, rather than seconds (see 'Firing four').

There is evidence to support each hypothesis — which might not be mutually exclusive — and they are the topic of much debate. Bulk endocytosis is the easiest to study, says Erik Jorgensen, a neuroscientist at the University of Utah in Salt Lake City. But full-collapse fusion is the one for which there is the most experimental support. Almost all the molecules involved in this process have been identified, says Rizzoli, although it is not yet clear how the clathrin-mediated process is regulated. And there are possible technical flaws in experiments that support the kiss-and-run hypothesis. "It has been named a biophysicist's fantasy by at least one very prominent synapse investigator," Rizzoli says.

Settling these differences is challenging — and each model may have a place. The number of molecules involved in the full-fusion process suggests that only around 10% of the vesicles at a synapse are releasing their cargo. Bulk endocytosis might be needed for the release of larger amounts of neurotransmitter.

But it is not clear when larger numbers would be needed. Rizzoli's team looked at neurotransmitter release in locusts that had been eaten by frogs but quickly removed from the frog's stomach¹. "Even under the extreme stress of being chased and eaten, the locust's synapses did not use more than 5% of their vesicles at one time," he says. "So in an *in vivo* context, the protein numbers of the clathrin pathway are completely sufficient."

TAG TRACKERS

Imaging can help to prove or disprove these hypotheses — and each type has both advantages and disadvantages. Electron microscopy has high resolution, but can work only on dead cells that have been chemically fixed. Fluorescence microscopy can image live cells, but will pick up components only if they can fluoresce under a particular wavelength of light — the rest of the neuron will remain dark. Researchers have been trying to get around these limitations by using dyes and protein analysis and by combining microscopy techniques.

Styryl dyes, for instance, can latch on to the membrane of the vesicle and thus help to image the vesicle as it travels in the neuron.

► and some even have a few million.

Researchers also know how these vesicles are called into action. When a neuron is activated, it fires an electrical pulse down its axon: a long fibre that behaves something like an electrical cable. When the pulse reaches the axon's tips, where most of the synapses lie, the vesicles there respond by moving to the synaptic membrane, merging with it and releasing their neurotransmitters. These messenger molecules then migrate to the neighbouring neuron across a gap called the synaptic cleft.

Much less clear is exactly what happens next.

"It is hard to get people to agree on much," says Silvio Rizzoli, a neurobiologist at the University Medical Center Göttingen in Germany. The vesicles must reform and refill with neurotransmitters very quickly, he says, otherwise the neurons would lose their ability to communicate and we would be paralysed in a matter of minutes. Working out how that happens has drawn researchers from a wide range of disciplines, from cellular and structural biology to physiology, physics and microscopy. "There is so much more to find out about how the single synapse operates," says electrophysiologist Ege

But they also stick to dying and dead cells, which clutters up the data, says Kavalali. As cells die, the lipids are moved around and this 'lipid scrambling' can lead to fluorescence not connected to vesicles.

Scientists have therefore tried to tweak the structure of the dyes so that they bind only to lipids in the vesicle membrane, with little success. Another approach is to genetically engineer the neuron's proteins to express tags that fluoresce under certain circumstances, creating 'reporter' molecules that can be used to trace what a vesicle is doing and where it is. SynaptopHluorin is a well-known reporter used in vesicles². It fluoresces green with a change in pH, which happens when the vesicle releases the neurotransmitter. Another pH-sensitive probe is pHTomato³, which shines red when the pH rises.

TECHNICOLOUR GLORY

The availability of probes that fluoresce in different colours lets scientists monitor multiple proteins in the vesicle and neuronal membrane. But there are challenges: it can be unclear whether the movement is coming from the vesicle or just the protein, says Kavalali.

To better understand synaptic dynamics, Rizzoli decided to identify the abundances and positions of neuronal proteins. He likens the pursuit to ecology research. A nineteenth-century scientist studying an area would

note the plants, the deer that eat the plants and the tigers that hunt the deer. Understanding the ecosystem in depth requires more than just counting the plants, deer and tigers, says Rizzoli. Knowing their locations and interactions is important as well. Similarly, some proteins and their functions in the synapse are known, but until we map the organization of their mini-jungle we will not know how they all function together, he says.

Around 20 years ago, Rizzoli's PhD adviser, physiologist William Betz, now at the University of Colorado in Aurora, developed dyes that can label living cells. But the cells needed to be chemically treated for the labels to identify the location of the proteins, and that made the dyes come off the membranes and get stuck elsewhere in the cell, resulting in poor quality images.

To try to find something better, Rizzoli and his team tested every marker available, including lipid tags, protein-coupled labels and simple stains. Nothing worked, he says. So he set out to synthesize a better one.

The result was membrane-binding fluorophore-cysteine-lysine-palmitoyl (mCLING)⁴. This marker clings to membranes and keeps clinging even after the neuron is treated with fixatives for imaging. In synapses, it labels mainly vesicles because there are few other membranes around, he says. Its chemical groups react more efficiently with fixatives than traditional probes and a long lipid tail anchors the probe in the membrane more strongly than the short tails of other commonly used probes.

Using mCLING, Rizzoli can track individual vesicles and see the neuron's

membrane fold inward and reform a vesicle after neurotransmitter release. Rizzoli has also scaled up his protein analysis. He and his team used data on protein types numbers and positions obtained from techniques such as staining, mass spectrometry, electron microscopy and super-resolution fluorescence microscopy (the technique responsible for this year's Nobel Prize in Chemistry), to build a three-dimensional computer model⁵

"There is so much more to find out about how the single synapse operates."

of an average synapse down to a resolution of 40 nm.

Rizzoli's team looked at the abundances of 62 different proteins in the synapse. The amounts differ wildly; there are 27,000 copies of one protein, for example, and 50 of another. "Together, they all sum to about 300,000 proteins in the model of the average synapse," he says. The researchers do not yet fully understand what they all do, but they hope that this grounding will set them on the right path.

One thing Rizzoli can say, is that the role a protein has in synaptic vesicle trafficking correlates strongly with its abundance in the neuron. The proteins involved in fusing the synaptic vesicle with the neuronal membrane are up to three levels of magnitude more abundant than those involved in the recovery of vesicles from the membrane after fusion.

This correlation can guide other researchers. A colleague recently contacted Rizzoli about a neuronal protein of unknown function. On the basis of its abundance, the pair concluded that it probably plays a part in the neuron's active zone, a region near the synaptic membrane where the neurotransmitter-filled vesicles wait to be called into action. But it is not plentiful enough to be involved in the retrieval of vesicle molecules, says Rizzoli, "which gives my colleague a nice place to start his experiments".

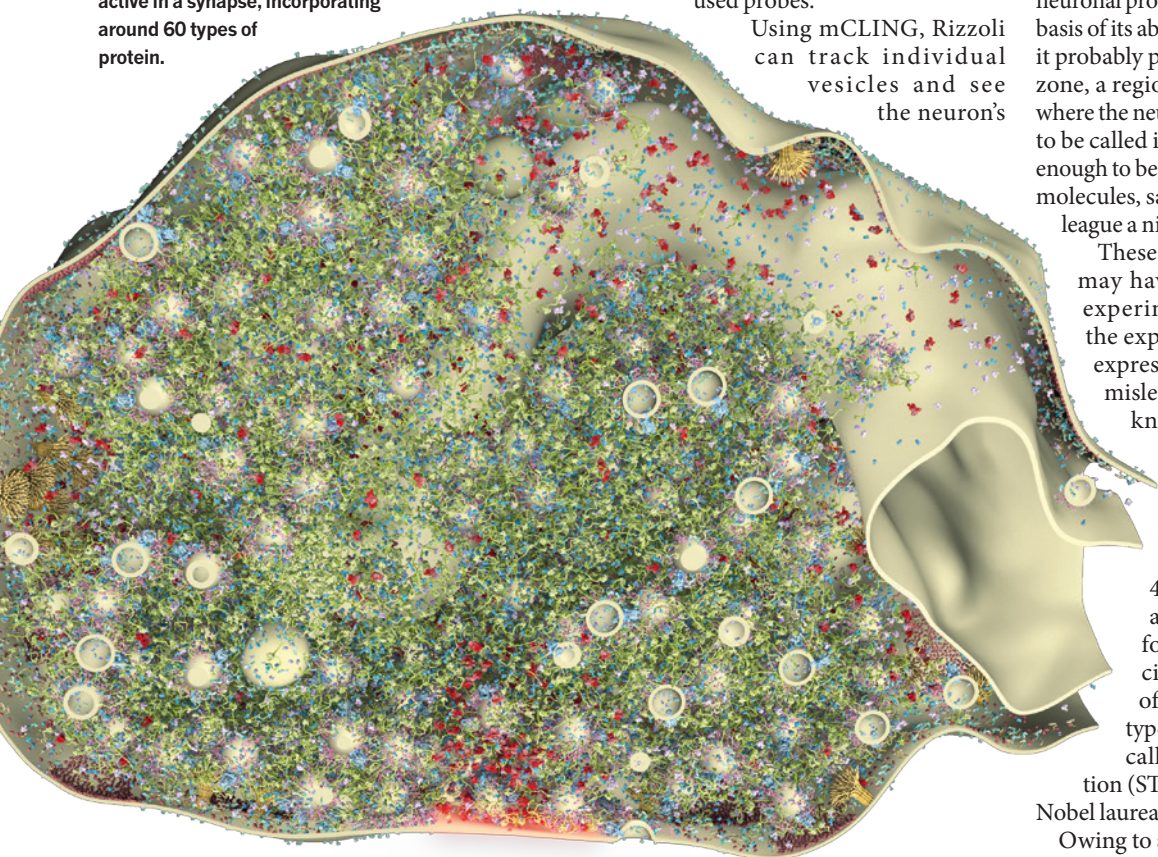
These data also suggest that scientists may have misstepped when they did experiments that involved boosting the expression of proteins. Such over-expression experiments "can be quite misleading", says Kavalali. "You don't know if what you are seeing are the genuine properties of this protein."

MORE MICROSCOPY

Rizzoli now wants to exceed the 40-nm resolution of his existing approach. He has been working for nearly a decade with physicist Stefan Hell at the University of Göttingen, who has invented a type of super-resolution microscopy called stimulated emission depletion (STED) and is one of the three 2014 Nobel laureates.

Owing to a fundamental limit imposed by

Around 300,000 molecules can be active in a synapse, incorporating around 60 types of protein.



S. RIZZOLI, UNIV. GÖTTINGEN MEDICAL CENTER

the physics of light waves and the aperture of the lenses, fluorescence microscopy cannot discern objects closer to one another than 200 nm, which is almost the size of the synapse. Super-resolution microscopy finds various ways around this limit.

For instance, STED starts with a laser that switches on fluorescent molecules in a 200-nm spot. A second laser then turns off all the molecules at the spot's edges. "The stronger the second laser, the more molecules it turns off," says Rizzoli, so the central area of fluorescence can be as small as 30 nm in a synapse. So, by scanning across a synapse one 30-nm region at a time, and measuring the fluorescence at each point, the STED system can build up an ultra-high-resolution image. "The procedure makes the initial blurry image far more clear," says Rizzoli (see 'Peek inside').

Rizzoli is also exploring isotopic microscopy, in which an ion beam plays the part of a light source in a traditional microscope and mass spectrometry detectors act as cameras. The instrument reconstructs an image of the sample as the beam burns off a sample's surface one atomic layer at a time then measures the atoms leaving the sample. "The resolution along the vertical axis is close to atomic size," he says. "So it really has some potential."

FLASH AND FREEZE

Jorgensen is also harnessing the high resolution of electron microscopy to home in on synaptic dynamics. In work published last December on neurotransmitter release in mouse neurons⁶, he and his team showed that the neuronal membrane folds inwards 50–100 milliseconds after the neuron is stimulated. This work led to the newest hypothesis about synaptic dynamics, ultrafast endocytosis. It suggests that new vesicle form immediately after the vesicle fuses with the membrane and releases the neurotransmitters.

For this work, Jorgensen and his team developed a technique nicknamed 'flash and freeze' electron microscopy. The researchers genetically engineer neurons so that a light beam causes the neuron to release neurotransmitters, then freeze different neurons at different time points between 15 milliseconds and 10 seconds after light stimulation and capture images of what is happening at the time.

Jorgensen says that his device expands a technique used by researchers Thomas Reese and John Heuser, who did some of the first work on electron-microscopy-based synaptic imaging⁷. The pair developed a freeze slammer, in which neurons were stimulated then thrown against a metal block that had been cooled with liquid helium, freezing their molecules in place.

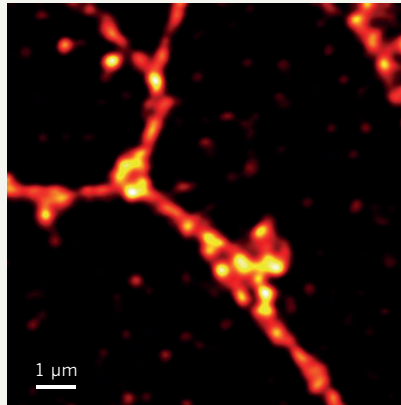
Jorgensen and his team also freeze neurons, but using pressure rather than temperature, which has the advantage of leaving more of the specimen free of ice crystals.

They also modified their

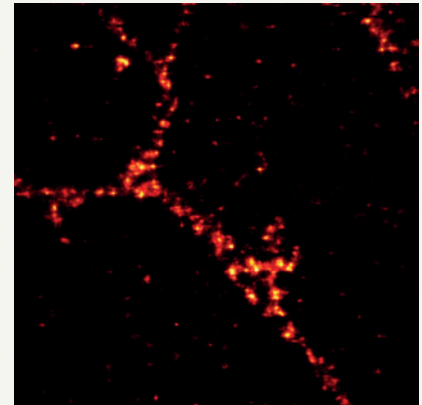
PEEK INSIDE

Owing to the physical properties of light, events in the synapse are too small-scale to discern with traditional confocal microscopy. By using fluorescent molecules and two lasers in sequence — a technique called stimulated emission depletion (STED) — scientists can capture some of the dynamic events in the synapse more clearly.

CONFOCAL MICROSCOPY



STED



electron-microscopy system, made by Leica Microsystems in Mannheim, Germany, to allow a path of light into the system's high-pressure freezers. Jorgensen says that they adapted the device out of necessity because "there is just no other way to look at fast events at the synapse".

Inspired by Jorgensen's idea, Leica is now building a light path into its new high-pressure freezers, known as EM HPM100, says Cveta Tomova, who manages electron-microscopy sample preparation products at Leica. The approach can help researchers to image cellular processes at millisecond and nanometre resolution and to avoid some issues with ice crystals, she says. The researchers can switch on a gene or process with light, then use high-pressure freezing to image the sample at that moment. Freeze-slamming is done at ambient pressure and allows thin samples to be frozen well up to a thickness of around 15 μm . High-pressure freezing means that researchers can avoid ice crystals down to around 200 μm under the surface.

Jorgensen is now working to tie in super-resolution microscopy techniques such as nano-resolution fluorescence electron microscopy (nanofEM) and super correlative light and electron microscopy (super CLEM), which can achieve resolutions of 20 nm. This gets close to the dimensions of proteins, which are 5–10 nm in diameter. His goal is to connect the ability to determine the location of proteins with the 1-nm resolution of electron microscopy. "We would then be able to characterize the molecular topography of the synapse," he says.

In his most recent work⁸, also using flash-and-freeze electron microscopy, he and his

team show how some of the hypotheses about synaptic dynamics might be linked. Using a technique called RNAi, in which small synthetic RNA molecules bind to — and block — the messenger RNAs that carry instructions for making proteins, they showed that the clathrin protein is an important component of ultrafast endocytosis, too. They also found that after neurotransmitter release, the bleb in the neuron's membrane separates completely to form what is known as an endosome. Synaptic vesicles then bud off from this endosome, rather than directly from the neuron's membrane. Rizzoli calls the new finding "bulk endocytosis on steroids".

What neuroscientists need, says Kavalali, is high-resolution images of living neurons stimulated to release neurotransmitters. He thinks highly of flash-and-freeze electron microscopy and has high hopes for super-resolution microscopy. These techniques will help researchers to map the synapse and track the events that occur before, during and after neurotransmitter release, and perhaps even to determine which hypothesis is the most accurate. ■

Vivien Marx is technology editor for *Nature* and *Nature Methods*.

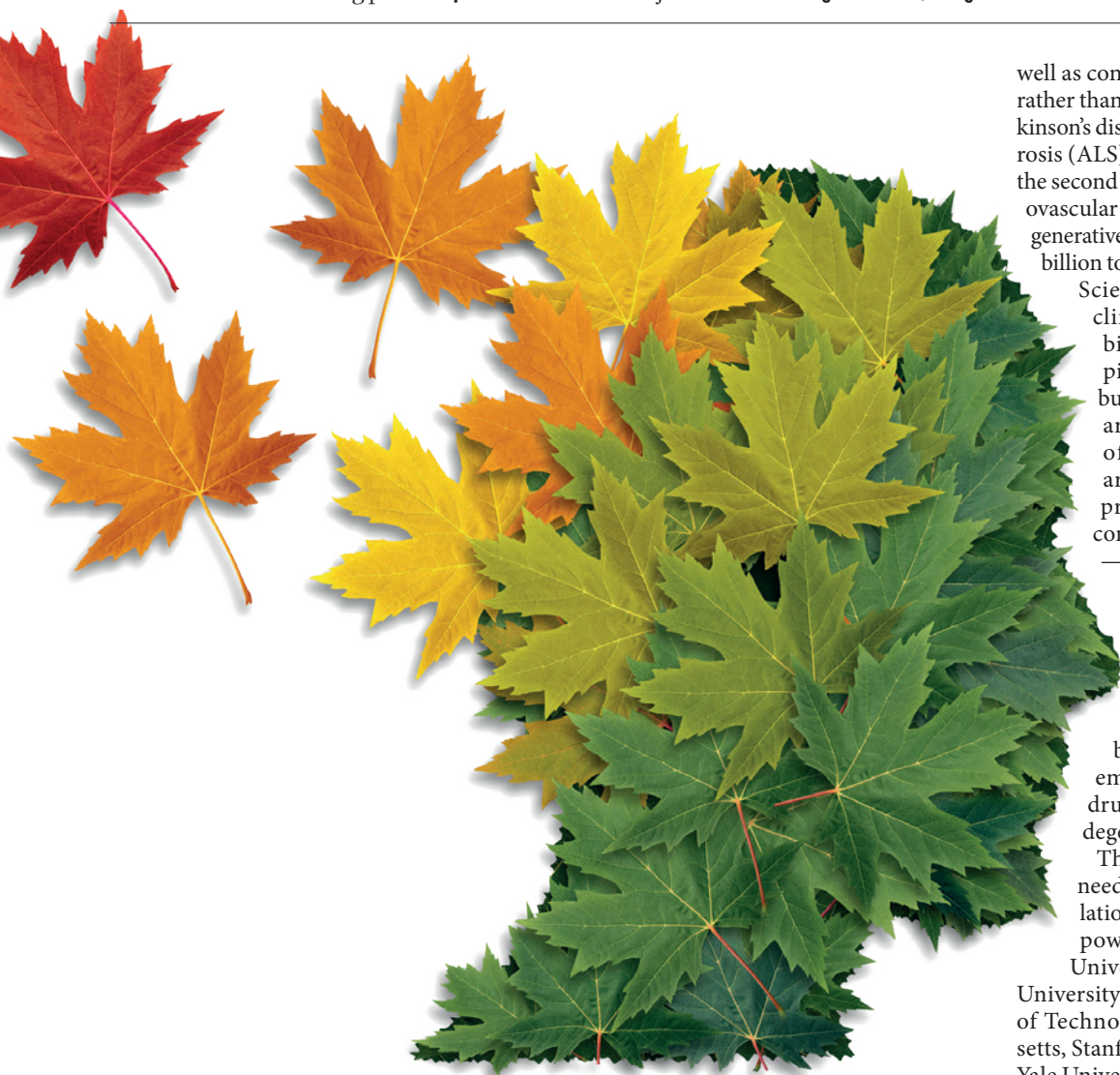
1. Denker, A. *et al. Proc. Natl Acad. Sci. USA* **108**, 17177–17182 (2011).
2. Miesenböck, G., De Angelis, D. A. & Rothman, J. E. *Nature* **394**, 192–195 (1998).
3. Li, Y. & Tsien, R. W. *Nature Neurosci.* **15**, 1047–1053 (2012).
4. Revelo, N. H. *et al. J. Cell Biol.* **205**, 591–606 (2014).
5. Wilhelm, B. G. *et al. Science* **344**, 1023–1028 (2014).
6. Watanabe, S. *et al. Nature* **504**, 242–247 (2013).
7. Heuser, J. E. & Reese, T. S. *J. Cell Biol.* **57**, 315–344 (1973).
8. Watanabe, S. *et al. Nature* <http://dx.doi.org/10.1038/nature13846> (2014).

CAREERS

TURNING POINT Cardiologist explains move from academia to big pharma **p.301**

@NATUREJOBS Follow us on Twitter for the latest on jobs and careers go.nature.com/e492gf

NATUREJOBS For the latest career listings and advice www.naturejobs.com



NEURODEGENERATIVE DISEASE

Brain windfall

Diseases such as Alzheimer's and Parkinson's are rising up the research agenda, partly because of ageing populations.

BY KATHARINE GAMMON

For those who study neurodegenerative diseases, there is a feeling of optimism — and trepidation — about what the next few decades will bring. A flood of dementia cases lies in humanity's future as the global population ages. The non-profit

group Alzheimer's Disease International predicts that there will be some 135 million cases worldwide by mid-century — up from 44 million cases today. And the World Health Organization (WHO) predicts that by 2040, as many developed countries' populations get older, neurodegenerative diseases such as Alzheimer's and other causes of dementia, as

well as conditions that affect mainly motor, rather than cognitive, functions, such as Parkinson's disease and amyotrophic lateral sclerosis (ALS), will overtake cancer to become the second leading cause of death after cardiovascular disease. Treatments for neurodegenerative diseases already contribute US\$20 billion to a market that is expected to grow.

Scientists who have both basic and clinical research skills in finding biomarkers and potential therapies could help to treat the rising burden of dementia. Governments around the world are well aware of the growing numbers, and are starting to invest in research programmes to understand and combat neurodegenerative diseases — which means that new research positions will emerge across the academic, non-profit and industry sectors worldwide.

Opportunities are growing in work to find biomarkers that can diagnose disease long before neurological symptoms emerge, and to identify targets for drugs that might halt or slow neurodegeneration.

Those who want to enter the field need to focus their research on translational targets. Training at one of the powerhouses in the field — such as University College London, Harvard University and the Massachusetts Institute of Technology in Cambridge, Massachusetts, Stanford University in California and Yale University in New Haven, Connecticut — puts candidates in a better position when they try to get a job, says Adrian Ivinson, director of translational research at the Harvard NeuroDiscovery Center in Boston, Massachusetts. "If you've got your training in a good place and you know how to translate it into a clinical setting, the future is looking good," he says. "The next few decades will change how we tackle these diseases." He adds that there is no magic bullet for a career path in the discipline, but advises that job-seekers would do well to find themselves in a vibrant lab environment, populated by both neuroscientists and clinicians. "Generally you would say that the richer and denser that commitment to talent, the better off you're going to be as a postgraduate in terms of your qualifications," he says.

Before neurodegenerative diseases can be ►

LIGHTSPRING / SHUTTERSTOCK

► treated, they need to be detected — and the earlier the better. As no reliable test yet exists to spot Alzheimer's disease, Parkinson's disease or ALS before symptoms appear, publicly funded agencies and pharmaceutical companies are pooling their resources to find biomarkers — arguably the area with the biggest career opportunities for researchers interested in entering the field. Biomarkers are measurable substances that can indicate early signs of disease, infection or exposure.

PUBLIC-PRIVATE INITIATIVES

Given this interest in early diagnosis and treatment for neurodegenerative diseases, job opportunities are starting to open up in academia and industry. In February, the US National Institutes of Health (NIH) launched a partnership with ten pharmaceutical companies, including Merck, Pfizer and Sanofi, and the trade organization Pharmaceutical Research and Manufacturers of America to find biomarkers and drug targets for Alzheimer's disease (along with type 2 diabetes and autoimmune disorders). The NIH and the drug-makers are combining resources and sharing information to make the process of drug discovery move faster — a type of partnership that has not yet been tried in this field. Some of the funds will go to three trials that will test drugs to slow the progression of Alzheimer's disease, which means job opportunities for young researchers.

Canada and Europe are following similar public-private models. The Canadian Consortium on Neurodegeneration in Aging was formed in September with Can\$55.5 million (US\$49.4 million) in funding from the Canadian government and other partners,

including Paris-based Sanofi. The consortium aims to develop ways to prevent, identify and slow the progression of neurodegenerative diseases. To do so, it is initially funding 47 principal investigators in 20 teams, across different universities, supporting about 340 scientists in all, including postdocs and graduate students. As the programme ramps up, the teams will be hiring young researchers from around the world.

In Europe, the European Union Joint Programme — Neurodegenerative Disease Research (JPND), headquartered in Paris, is the largest neurodegeneration-research initiative worldwide, involving the 29 member and third-country member countries. The consortium is funding studies on early-onset dementia, and on biomarkers and gender differences in Alzheimer's and other diseases. It funds basic-science research through to health and clinical care — which means that jobs are now opening up at institutions funded by the JPND. The idea is that no single country can have the breadth and depth of research needed to mitigate the coming avalanche of dementia cases, so teaming up is the best option.

Industry is also jumping on board. Biogen Idec in Cambridge, Massachusetts, and Genentech in South San Francisco, California, are recruiting discovery and translational scientists in neurodegenerative research as well as medical directors in development, experimental medicine and biomarkers, and global medical affairs. In the United Kingdom, GlaxoSmithKline has recently joined the UK Dementias Research Platform, a multimillion-pound public-private partnership developed and led by the UK Medical Research Council, which is focused on accelerating progress in neurodegeneration research. The platform is looking to validate biomarkers for different stages of disease progression.

Other approaches to tracking down the roots of neurodegeneration focus on neuronal synapses — the connections between brain cells. In October, the non-profit Allen Institute in Seattle, Washington, was awarded an \$8.7-million, five-year grant to create a model of all the genes and proteins involved in different synapse populations in both the human and mouse brain that will be accessible to all researchers. The project aims to gain an understanding of how individual neurons in the brain connect and transfer information through synapses, says spokesman Rob Piercy. Many brain diseases and disorders ranging from Parkinson's disease to depression are rooted in abnormalities of synaptic function, he says. In the next few years, the institute will be nearly doubling its payroll from 270 to 500, and one-fifth of those hired will be brain-disease researchers.

The field has become extremely multidisciplinary, says neurosurgeon William Mack, director of the cerebrovascular laboratory in the Zilkha Neurogenetic Institute at

the University of Southern California in Los Angeles. Those trained in imaging, radiology and computing have a strong chance of getting hired. Teaching hospitals are also assuming a leading role in neurodegeneration research, adds Iverson, because they can bring basic-research findings to the clinic. "You'll see a lot of funding attention go to the groups that can demonstrate they're ready to pivot from basic mechanisms to practical questions they can address," he says.

Joseph Mazzulli, a neuroscientist at Northwestern University's Feinberg School of Medicine in Chicago, Illinois, is linking his basic research with industry to bring a potential neurodegenerative treatment to market. In 2011, he co-founded Lysosomal Therapeutics, a start-up in Cambridge, Massachusetts, that aims to market drugs to treat Parkinson's disease and Gaucher's disease, which leads to motor impairments similar to parkinsonism. He and his partners were approached by drug-makers about potential collaborations, but they chose instead to launch their own start-up to test compounds that target glucocerebrosidase, an enzyme that is involved in both Parkinson's and Gaucher's diseases. "We felt that we had a firm grasp on the mechanism that linked these two diseases," he says. "Ultimately, we thought doing it ourselves would accelerate the process and get a drug to patients more quickly." Investors seem to like the concept: the fledgling company raised \$4.8 million in its first round of funding in May 2014.

MEETING OF WORLDS

Balancing research and the clinic may seem like a stretch, but for neurologist Vikram Khurana, it comes naturally. With posts at both Massachusetts General Hospital in Boston and the Whitehead Institute in Cambridge, he is studying balance disorders. Last year, Khurana and colleagues discovered a compound that can both identify and reverse some of the damage that occurs in the brain cells of patients with Parkinson's disease. He may launch a start-up to commercialize the discovery. He sees juggling his clinical and research careers as his future in the field of neurodegeneration. "It's a 'jack of all trades' career path for sure," he says, "but I recommend it to anyone who thrives on bringing different worlds together."

Iverson adds that the field has changed since he started his career a decade ago. "We now have such an abundance of good questions to ask, be they clinical trials or fairly basic neuroscience questions or translational projects in between," he says. "Compared to when I got into this field, I'd say there's real energy that will also fuel job prospects." ■

Katharine Gammon is a freelance writer in Santa Monica, California.



Vikram Khurana is working to reverse the brain degeneration associated with Parkinson's disease.

THE WHITEHEAD INSTITUTE

TURNING POINT

Morris Birnbaum



Morris Birnbaum, an academic cardiovascular researcher for more than 20 years, left the University of Pennsylvania (Penn) in Philadelphia in May this year to become chief scientific officer for cardiovascular and metabolic disease research at pharmaceutical company Pfizer in New York City. He explains how the move has been both daunting and invigorating.

What was your postgraduate training like?

I always knew that I liked research and wanted to do it. After medical school, where I trained in internal medicine, I thought I would have an active clinical and research life. The molecular-biology revolution was just starting and I decided to go to the University of California, San Francisco, as a postdoc to get skills in that discipline. It was a complete disaster.

What went wrong?

Due to my clinical training, I had not been in a lab for five years. I deliberately picked that discipline because it was completely different from what I did as a medical student — I knew it was going to be important to have a grounding in molecular biology and genomics. But as a result, I knew absolutely nothing about it, technically or intellectually. I also was not prepared for the level of independence that I was given in a large lab. I floundered there — I did not know what a good project was. I never got anything to work well and never got a publication.

How did you recover?

The smartest decision I ever made was realizing that I was not ready to run a lab. I turned

down job offers to be an assistant professor and took a second postdoc in molecular biology at Memorial Sloan Kettering Cancer Center in New York City. I realized that I had control over my life and scientific productivity. It's interesting. Two messages that I try to get across to trainees are: you do not have to have success quickly to have a great career, and the only person you can depend on is yourself.

Why did you never practise medicine?

After my second postdoc, I took a job as an assistant professor at Harvard Medical School in Boston, Massachusetts, thinking that the moment I had my lab up and running, I would start seeing patients. As the years went by, I realized that for me, basic research was a full-time job. Even though I trained as a physician, I never saw a patient again. I do not regret the medical training, though. It was a major time of personal development and it has informed my science ever since. I have always run my labs to focus on projects with a physiological relevance. I stayed at Harvard for roughly seven years and then moved to Penn. There I secured a Howard Hughes Medical Institute appointment (which provides full salary and benefits, and a research budget for five years) to study glucose transport.

How did your research focus shift?

When I cloned the gene *GLUT4*, which specifies the genetic code for an insulin-regulated protein that facilitates glucose transport into and out of tissues, it changed the field of insulin action. It allowed us to focus on the mechanism of action of insulin in muscle and fat. I had to retool completely. I basically became a cell biologist. By the time I moved to Penn, I was frustrated — it had been hard to do *in vivo* physiology at Harvard because of a lack of access to animal models. But after the move, I started using mouse models. For 20 years, my lab studied aspects of how insulin works and how the body deals with nutrients after a meal.

It sounds like you were having a good run. Why move to Pfizer?

I was arguably at my most productive, doing exciting stuff. But I had been on the advisory board at Pfizer for ten years. I had some knowledge of what was happening at the company. I agreed to talk to Pfizer because my wife was interested in moving closer to family in Boston, but I made it clear that the chances of my moving were vanishingly small. As I continued the interview process, however, I kept asking myself whether to stay or go. The decision was

down to which path would prove more challenging. I realized that staying in academia, even though it is harder now to secure funding and publications, was the easier route. Trying something that I knew nothing about — that could have a big impact — was much harder. It got me excited about the possibilities.

Did changes in academia affect your decision?

Somewhat. One that did is the increasing importance placed on publications. Twenty years ago, a scientist was judged on a body of work; people worked on the same thing for long periods of time. In many labs now, the goal of doing science is the publication.

What has been the biggest adjustment in working at Pfizer?

Two aspects have been difficult. First, the size and scope of the job are much larger now; for example, I have to stay up not just on the science of developing a drug, which includes chemistry, pharmacology and drug safety, but also on the business, marketing and competitive-landscape side. There is also a hierarchical part that is different from academia, where the boss concept does not really exist.

Any unexpected surprises?

Yes. One of my big fears about the move to Pfizer was that instead of focusing on science, I would be stuck in meetings all day. That is not the case. I do the same fundamental exercise each day — interpreting data and doing experiments.

Can you dispel any myths about the pharmaceutical sector for young researchers?

Yes. It's a myth that you cannot do interesting science in industry and that you have no independence. For example, our postdocs work on non-drug-related basic investigations so there are no restrictions on publishing their work. The other myth is that academia is collaborative and the pharmaceutical sector is not. Nothing could be further from the truth. Here, people are judged much more by how they contribute to a successful project.

Do you think you will look back on this as a good career move?

The philosophy here is, 'Give leaders autonomy and hold them accountable.' As the months roll by, it is becoming more and more clear that I am going to have to establish successful drug-development programmes. The goals in drug discovery are different from those in academia and impossible to fake: get something on the market to help somebody. But I am already feeling how much pressure that places on us to succeed at something that is really hard. ■

INTERVIEW BY VIRGINIA GEWIN

MÖBIUS

Repeat after me.

BY J. W. ALDEN

I've watched you die a thousand times. The first was the hardest. I watched the birds from my hotel balcony, tracing every arc and dive, marking each wire they touched, each windowsill. If one of them hadn't landed on the stoplight above you, I might never have noticed you crossing the street. I might never have noticed the van, a blurry streak of silver barrelling your way.

Your first death was the hardest, but it was the most beautiful. I had to see it again.

Chaos touches those around you, extending outwards like ripples in a pond. The birds scatter into the air. A man walking his dog on the sidewalk drops the leash. A woman three paces behind you stumbles backwards with hands over her mouth. Teenagers huddled around a nearby storefront reach for cellphones, one to dial 911, the others to snap pictures.

The look in your eyes never changes, that sudden dilation just before fluttering to a close. They shimmer like starlight, even as your head snaps back and auburn wisps tangle in the windshield wipers.

The driver's name is Will. He's 42. He has a wife and a daughter. He told me over a cup of coffee at the bar he'd been in that afternoon. He seemed nice. Not the killing type, I think. But a conversation won't change what happens, no matter how many times we have it. This last time, we talked about God and destiny and drunk driving, just before the quantum beacon beneath my ribcage came to life and brought me to the beginning again, as it always does.

It's not Will's fault. He's a slave to this moment, entangled. Like you. Like me.

They told me that this could happen. Over the years, I heard stories of beacons blinking out, operatives lost. Never the whos, whens or whys, only that it's happened before and will happen again. I shrugged away those warnings, certain I'd never get caught up

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

in extracurricular voyeurism. Tracking the H7N9 mutation was too important. Or so I believed then.

But the mission grows stale. I'm weary of chasing birds, watching money change hands. I observe the same moments again and again, unable to do anything but watch, learn, report. Just once, I want change. *Permanent* change that won't be erased when the beacon chimes and the machines bring me home for a pat on the head — or back



to the start for another try. They keep saying the government will let us *prevent* the outbreak one day. But we've been at this for years, and it's clear we're just going to keep watching. We're not the heroes they said we'd be. We're just filling in the history books.

I don't know what it is about you, your death, your moment. But the first time I called out and you looked up as the silver streak took you, pointed your chin in a slightly different direction from the previous hundred times, I knew I'd never free myself of this.

I tried to learn your name once. You scowled, unaware of the favour I'd done you. Silver drifted by, your fragile form unassailed. It didn't matter. My interval expired, and the beacon chimed. I woke again in that hotel, across from that intersection, ready to watch you die or not die for the thousandth time.

It will go on like this until I can accept the futility of it. I can weigh your fate again and

again, or I can turn away, carry out my task in your time, and make my report — leaving you, Will and all the others forever at the mercy of this moment.

I've seen it from every angle now. Each time, I come closer to doing what must be done. Each time, I wonder how long they'll let me do this. They must have drilled this scenario. There must be some contingency. Maybe the next time my beacon goes off, I'll wake with a bright light in my face, people in lab coats and army uniforms asking how many tries it takes to find one lousy bird. Or maybe I'll just keep going, on and on, until I decide I'm done.

At the very least, my suspicions have been confirmed, one death at a time. It's not about you. Never was. We cling to each heartbeat, each fleeting moment, desperate to make use before they're gone. But I've found one I can have and hold, one I can love and abuse and make my own. Maybe even one I can change.

You? You just happen to be there for it.

I'm not sure why you smile this time. Maybe it's the way the sun hits my shoulders. Maybe it's the bounce in my step as I cross the street.

For the first time since the first time, I realize how unaware you are of this moment we're caught in. You don't know about Will. You don't know about the dropped dog leash or the camera phones. You don't know that I've saved your life again and again, that I've let you die a thousand times. You also don't know how the beacon works or what a piezoelectric power source is. You don't know that it will cease to function when my heart stops beating.

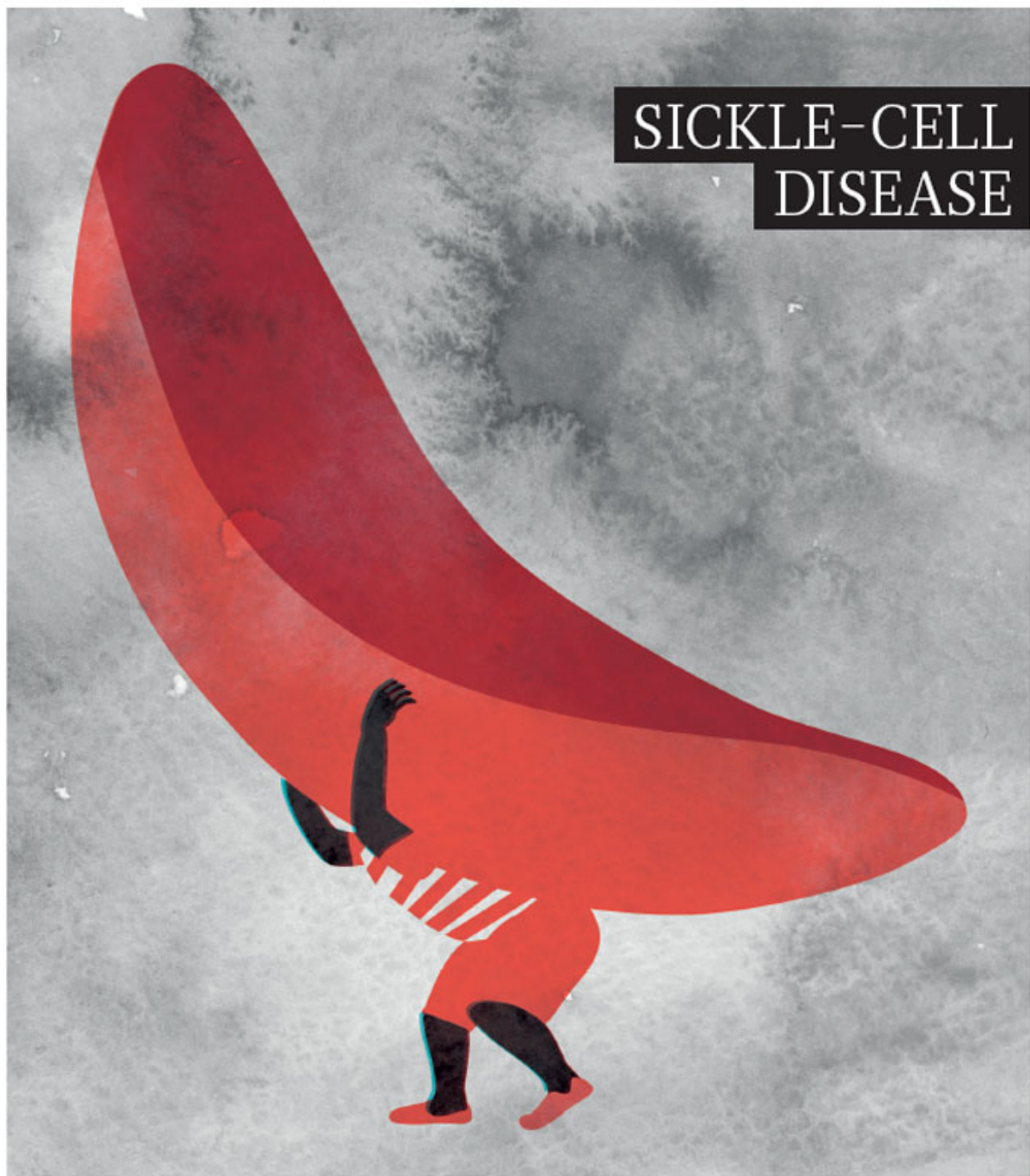
Your smile fades when I wrap my hands around your shoulders. A trembled shout leaves your lips as I shove you backwards onto the sidewalk. I mouth a silent apology for robbing you of this moment as I step into the path of the van. Your eyes grow wide, but it's not the starry shimmer I notice this time; it's the silver streak within. ■

J. W. Alden lives near West Palm Beach, Florida with his wife, Allison. He's a graduate of Odyssey Writing Workshop and a member of Codex Writers. Find him at www.AuthorAlden.com.

JACEY

natureOUTLOOK

SICKLE-CELL DISEASE



Produced with support from:



Steps to ease
the burden

natureOUTLOOK

SICKLE-CELL DISEASE

13 November 2014 / Vol 515 / Issue No 7526



Cover art: André da Loba

Editorial

Herb Brody,
Michelle Grayson,
Lauren Gravitz,
Stephen Pincock,
Kathryn Miller

Art & Design

Wesley Fernandes,
Mohamed Ashour,
Kieran McCann,
Kate Duncan

Production

Karl Smart,
Ian Pope,
Robert Sullivan

Sponsorship

David Bagshaw,
Yvette Smith

Marketing

Hannah Phipps

Project Manager

Anastasia Panoutsou

Art Director

Kelly Buckheit Krause

Publisher

Richard Hughes

Chief Magazine Editor

Rosie Mestel

Editor-in-Chief

Philip Campbell

The fate of people with sickle-cell disease is largely determined before they are born. Those in developed nations can be expected to live into their 40s, 50s and beyond. In low-income countries, including many African nations, up to 90% of people with the genetic blood disorder die before their fifth birthday (page S2). This stark difference is attributed to newer, life-long interventions that begin at birth — ones that only those in rich countries can afford.

But even patients with access to good medical care still suffer from painful and debilitating symptoms. Only one drug has been approved to treat the disease (S4), and therapies for the pain crises that accompany sickle cell are limited because researchers are only beginning to understand the underlying mechanisms of the agony (S8). Some serious manifestations of the disease, stroke in particular, can have a permanent effect on both physical and intellectual well-being (S16).

Treatments and technologies offer tantalizing hopes of a cure for those who can afford them. Stem-cell transplantation, which has been used to reverse sickle-cell disease in hundreds of children, is being adapted so that adults can benefit (S14). And the advance of gene-editing techniques could soon allow scientists to perform molecular microsurgery, repairing the genetic error that causes the disorder (S11).

Overcoming the global inequity in sickle-cell outcomes will require difficult but feasible steps involving diagnosis, drug availability and better awareness of the disease among primary-care providers (S10). Testing and education programmes may help to improve treatment and prevent the spread of disease in developing countries (S6).

We are pleased to acknowledge the financial support of the Doris Duke Charitable Foundation and Emmaus Life Sciences, Inc. in producing this Outlook. As always, *Nature* retains sole responsibility for all editorial content.

Lauren Gravitz and Stephen Pincock
Contributing Editors

CONTENTS

S2 EPIDEMIOLOGY

A moving target

The link between sickle-cell disease and geographical location

S4 DRUG DEVELOPMENT

A complicated path

Efforts to increase the number of drugs available to treat sickle-cell disease

S6 Q&A

Healthy progress

Physician Graham Serjeant explains the significance of his work in Jamaica

S8 NEUROBIOLOGY

Life beyond the pain

Researchers are looking at ways to ease the agony caused by sickle-cell disease

S10 PERSPECTIVE

We need a global solution

Isaac Odame argues the case to end discrepancies in medical treatment

S11 GENE THERAPY

Editorial control

Gene editing is a promising method to fix the disease's minor genetic error

S14 STEM CELLS

Creating a cure-all

Efforts to increase availability of stem-cell transplantation

S16 PERSPECTIVE

Thinking beyond survival

We need to consider childrens' quality of life, says Michael R. DeBaun

COLLECTION

S17 Lysine-specific demethylase 1 is a therapeutic target for fetal hemoglobin induction
Shi, L. et al.

S21 Allogeneic cellular and autologous stem cell therapy for sickle cell disease: 'whom, when and how'
Freed, J. et al.

S31 Emerging science of hydroxyurea therapy for pediatric sickle cell disease
Green, N. S. & Barral, S.

S40 Genome-wide association study identifies genetic variants influencing F-cell levels in sickle-cell patients
Bhatnager, P. et al.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at go.nature.com/e4dwzvw

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol. XXX, No. XXXX Suppl., Sxx–Sxx (2014).

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Sickle-Cell Disease* supplement can be found at <http://www.nature.com/nature/outlook/sickle-cell>. It features all newly commissioned content as well as a selection of relevant previously published material.

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group – Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES

Feedback@nature.com
Copyright © 2014 Nature Publishing Group

A moving target

Migration has led to an increase in the occurrence of sickle-cell disease in countries with previously low incidence of the disorder.

BY SIMON PLEASANTS

Sickle-cell disease is a heritable blood disorder that can have devastating effects. It is the result of a minor genetic change — a single error in a single gene — yet it is one of the most common genetic causes of sickness and death¹. It is also on the rise. According to a study published last year, approximately 305,800 children were born with sickle-cell disease in 2010, about two thirds of them in Africa². By 2050, the study authors predict, those numbers will increase by more than 25%, to more than 400,000 babies born annually with the disorder.

The disease-causing mutation occurs in a gene responsible for the production of haemoglobin, the protein in red blood cells that carries oxygen to tissues throughout the body. Individuals who inherit only one mutant gene typically experience no ill effects, but those who have two faulty copies develop the full-blown disease. The mutation itself is tiny: it replaces one amino acid (called glutamic acid) with another (valine) at one location in the haemoglobin protein. This small change has a major consequence — it turns the flexible, disc-shaped red blood cells rigid, giving them a distorted, sickle-like appearance.

Sickled cells are not only less pliable than healthy red blood cells, they're also stickier. This causes them to adhere to each other and plug up narrow blood vessels, reducing blood flow and preventing adequate oxygen from being delivered to the tissues that need it. This in turn leads to many complications, some of which are life threatening. People with sickle-cell disease commonly experience episodes of acute pain, known as crises, that can last from a few hours to a few days. Other complications include anaemia, leg ulcers, jaundice, kidney damage, high blood pressure, gallstones, increased susceptibility to infections and stroke. These problems arise from a variety of complex mechanisms, such as altered pain receptors (see page S8) and the decreased lifespan of red blood cells¹.

In high-income countries, improved management of the disease and medication, including strong painkillers, has meant that patients with sickle-cell disease can expect a better quality of life. In low-income nations, however, those with sickle cell are not as fortunate: in Africa, at least 50% of children born with the condition die before their fifth birthday³.

Sickle-cell disease is most prevalent in areas with hot climates — Africa, the Indian

subcontinent and the Caribbean are all affected. There is evidence to suggest that the geographic concentration of people with sickle-cell disease arose because people with sickle-cell trait — those who carry the sickle-cell gene but are generally healthy — have enhanced resistance to malaria⁴, resulting in preferential selection of the gene.

The malaria-causing parasite typically invades red blood cells and multiplies inside them. But although sickle-cell carriers can be infected, the presence of sickle haemoglobin confers resistance to malaria. There are a number of different theories, one of which is that subtle biological changes in the blood — pos-

of Africa — a crippling health burden in a region that is ill-equipped to bear it, and one that lacks the financial and health-care resources present in wealthy countries.

ON THE MOVE

Increased population movement associated with globalization is bringing sickle cell to regions that previously had negligibly low incidences of the disease. Migration of people with the sickle-cell trait or sickle-cell disease caused the faulty gene to flow from high-frequency sickle-cell gene areas, such as Africa and India, to western Europe, the eastern coast of South America, and North America⁷. These migra-

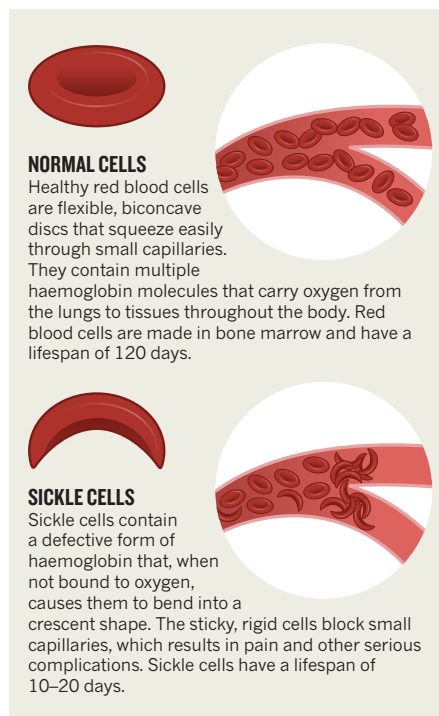
tions have increased disease prevalence in areas not previously associated with the disorder. The Sickle Cell Disease Association of America estimates that there are about 100,000 people living with the disorder in the country. The US

“In Africa, at least 50% of children born with sickle-cell disease die before their fifth birthday.”

Department of Health and Human Service's National Heart, Lung and Blood Institute suggests that sickle-cell disease affects about 1 in 500 African American babies, a number about ten times greater the general population.

Ironically, improved medical treatment leading to longer life expectancy is contributing to the increased prevalence of sickle-cell disease. Penicillin, taken daily, helps to prevent and treat infection in young children, folic acid stimulates the production of red blood cells, the drug hydroxycarbamide (known as hydroxyurea in the United States) reduces the number and severity of painful crises, and blood transfusions help to temper acute cases of the disease. With so many interventions, more people will survive to pass on sickle-cell genes to their children. ■

Simon Pleasants is a copy editor at Nature.



sibly higher carbon monoxide levels⁵ or the interaction between oxygen and sickle haemoglobin⁶ — makes people with the sickle-cell trait less of a hospitable host to the parasite and usually prevents the disease from taking hold.

About 70% of global cases of sickle-cell disease are in Africa. The World Health Organization estimates that in some African countries, including the Democratic Republic of Congo, Ghana and Nigeria, prevalence of sickle-cell trait is between 10% and 30%, resulting in an estimated disease prevalence of at least 2% in most

1. Makani, J. *et al. Sci. World J.* <http://dx.doi.org/10.1155/2013/193252> (2013).
2. Piel, F. B. *et al. PLoS Med.* <http://dx.doi.org/10.1371/journal.pmed.1001484> (2013).
3. Grosse, S. D. *et al. Am. J. Prev. Med.* **41**, S398–S405 (2011).
4. Allison, A. C. *Br. Med. J.* **1**, 290–294 (1954).
5. Ferreira, A. *et al. Cell* **145**, 398–409 (2011).
6. Cyrklaff, M. *et al. Science* **334**, 1283–1286 (2011).
7. Piel, F. B. *et al. Lancet* **381**, 142–151 (2013).

CARRIED IN THE BLOOD

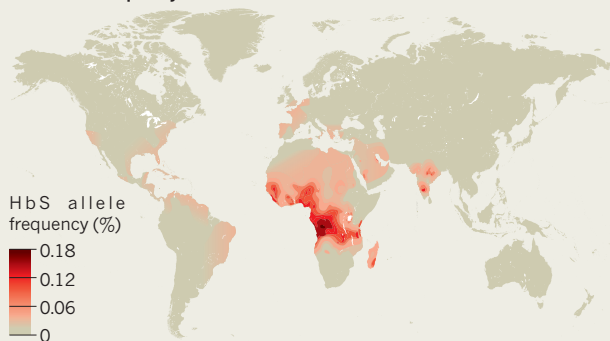
Sickle-cell disease is often fatal in low-income nations. But researchers are beginning to understand more about the disease and how to treat it. By Simon Pleasants.

FATED LINK

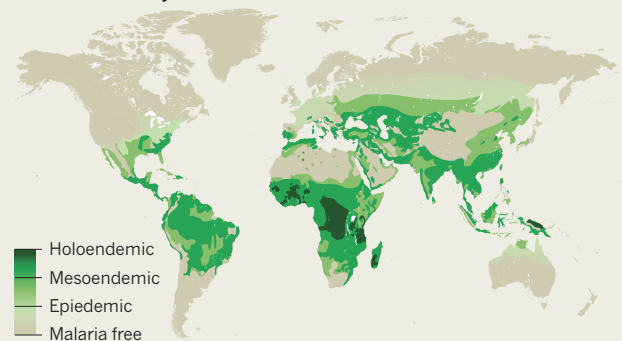
It is no coincidence that sickle-cell disease and malaria share similar global distributions. The gene responsible for sickle haemoglobin confers increased resistance to the malaria parasite when inherited from only one parent. Such a trait allowed carriers of the gene (HbS distribution, on the map below in red)¹ to survive and reproduce in areas where malaria is present (green)².

Although malaria is indigenous in many areas, the sickle haemoglobin gene has a more narrow distribution than malaria because the mutation arose in Africa. Limited migration from Africa to certain regions (for example, China) has restricted its spread.

HbS allele frequency

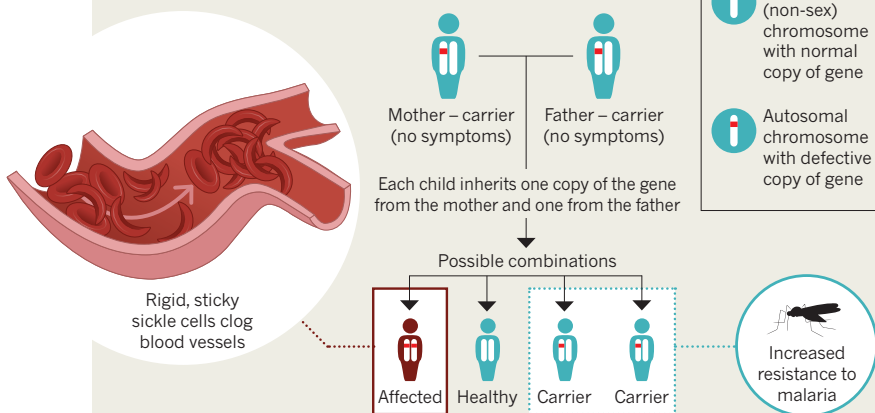


Malaria endemicity



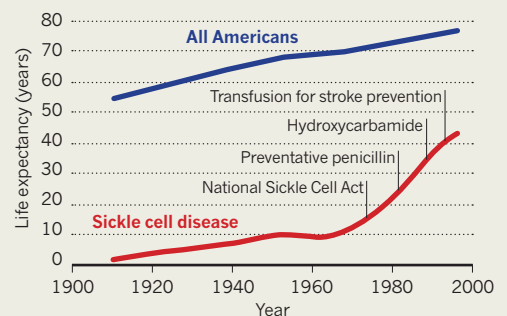
GENETIC ROULETTE

If both parents are carriers of the sickle haemoglobin gene, each child has a one-in-four chance of inheriting the full-blown disease.



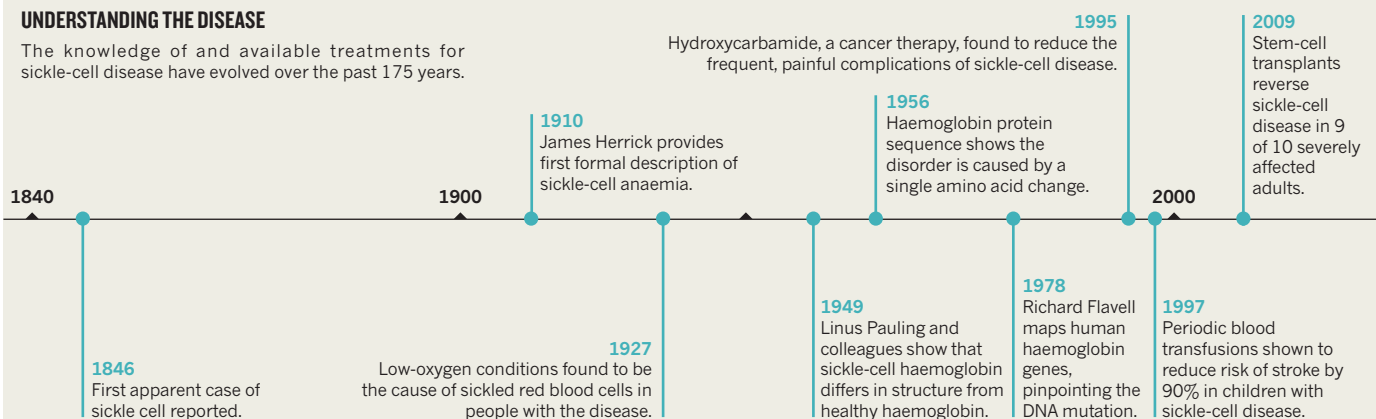
HEALTH IMPROVEMENT

As a result of more effective treatments, the life expectancy of people with sickle-cell disease in the United States has increased dramatically. Most developing nations, however, have yet to see such a change³.



UNDERSTANDING THE DISEASE

The knowledge of and available treatments for sickle-cell disease have evolved over the past 175 years.



References: 1. Piel, F. B. *et al. Lancet* **381**, 142–151 (2013); 2. Piel, F. B. *et al. Nature Comms.* **1**, 104 (2010); 3. US National Institutes of Health: National Heart, Lung and Blood Institute, available at: go.nature.com/4cuwgj.



ANDRÉ DA LOBA

DRUG DEVELOPMENT

A complicated path

Only one drug is available to treat sickle-cell disease, but a wave of investment and industry attention is set to turn the tide.

BY COURTNEY HUMPHRIES

Sickle-cell disease makes a sweeping hit to the body. What begins as one small genetic alteration — a single faulty nucleotide in a haemoglobin gene — ends in a cascade of damage. People with sickle-cell disease have a mutant form of haemoglobin, the molecule inside our red blood cells that allows them to carry oxygen to tissues throughout the body. This form, called 'sickle' haemoglobin, is capable of carrying oxygen, but once the oxygen is released the haemoglobin joins with other sickle haemoglobin molecules to create rigid rods that bend the normally flexible, disc-shaped red blood cells into distorted, narrow crescents. The sickled cells are twisted, dense and dehydrated, and clog blood flow when they stick to the walls of blood vessels and to other blood cells. Over time, the consequences accumulate: anaemia, periods of overwhelming pain, greater risk of infection, damaged organs and tissues, and a shortened lifespan.

Treating the disorder has proved to be far more complicated than describing it. Drug development has been hampered by both limited investment from the pharmaceutical industry and limited trial enrolment, leaving patients

with few options. In addition, the disease is rare in the US and Europe, which are responsible for the majority of clinical trials — there are millions of cases of the disease worldwide, yet only 100,000 people living in the United States have it. Researchers have a detailed scientific understanding of the condition, but there is only one drug available to treat sickle-cell disease: hydroxycarbamide (known as hydroxyurea in the United States), approved by the US Food and Drug Administration in 1998. And although it can reduce the number of painful episodes and hospitalizations by preventing blood vessels from clogging, only about two-thirds of adult patients have any response.

Bone-marrow transplants can provide a cure because they replace the faulty haematopoietic stem cells that produce sickled cells with ones that make healthy red blood cells (see page S14). But the procedure is expensive, has risks such as infections and life-threatening immune-system reactions, and is not yet widely available. Other treatments have been slow to emerge, but there are reasons to be optimistic. "A lot of new drugs are being pursued as specific agents for sickle cell," says Betty Pace, a paediatric haematologist and oncologist at Georgia Regents University in Augusta. Pharmaceutical companies are

increasingly seeing potential for profit in treatments for diseases that are rare in the developed world. Therapies are being developed to prevent complications such as stroke or organ damage or to treat complications as soon as they arise. There are more clinical trials and research has uncovered potential disease targets that could lead to new therapies.

HAMMERING THE SICKLE

One sign of the growing corporate interest in sickle-cell disease occurred in July, when pharmaceutical and medical-device company Baxter International, headquartered in Deerfield, Illinois, acquired a potential drug called Aes-103, together with its parent company, AesRx. It took more than a decade to secure investment for Aes-103, which illustrates the challenges that scientists face when developing treatments for rare diseases.

Don Abraham, a chemist at Virginia Commonwealth University in Richmond, had been working since 1975 to develop a drug that binds to sickle haemoglobin and prevents it from distorting red blood cells into the characteristic crescent. His problem was one of scale. Patients with sickle cell have around 320 grams of haemoglobin in their bodies, and most of the

compounds he tried were too toxic at the doses needed to make a difference. Thinking that natural food components might be less toxic, Abraham began looking for safer solutions in the human diet. In 2000, his search for anti-sickling food agents led him to a naturally occurring chemical called 5-(hydroxymethyl)furfural, or 5-HMF, which is found in caramel, roasted coffee and dark beer. After decades of testing agents in his laboratory, says the now-retired Abraham, “it is by far the best thing we’ve seen.” But in spite of promising results in animal studies¹, which showed that the molecule safely prevented red blood cells from sickling in mice engineered to have the gene for human sickle haemoglobin, it experienced a series of setbacks.

The drug was initially licensed to a now-defunct pharmaceutical company but “was sitting on a shelf”, says Stephen Seiler, who formed biopharmaceutical company AesRx of Newton, Massachusetts, in 2008 so that he could buy the rights to develop the compound as Aes-103. When venture capitalists showed little interest in backing the work, Seiler went to the US government and received funding as part of an unusual public–private partnership.

As an anti-sickling agent, Aes-103 represents “a class of drugs that’s been dreamed about for as long as we’ve thought about treatments for sickle-cell disease”, says Gregory Kato, director of the Adult Sickle Cell Disease Center of Excellence at the University of Pittsburgh in Pennsylvania. But so far, it is the only such drug to enter clinical trials. AesRx worked with Kato, who was at the US National Heart, Lung, and Blood Institute in Bethesda, Maryland, at the time, and with researchers at the National Institutes of Health’s Therapeutics for Rare and Neglected Diseases programme to take the drug through early-stage clinical trials. In a phase I/IIa safety study, 15 patients who took one dose of Aes-103 experienced significantly less pain than patients given placebo. The drug is now in a phase II trial — designed to test dosing and efficacy — in London, which is due for completion in 2015.

When Seiler first dreamed of bringing Aes-103 into clinical trials, he was one of a very few working on sickle-cell treatments. But interest has grown substantially. “Orphan drugs are a lot more attractive than people perceived them to be in 2008,” he says. Because orphan diseases are uncommon, regulatory standards are more flexible to encourage companies to invest in them, so companies can run smaller, less expensive trials, which are often accompanied by government incentives, such as a US tax credit on the costs of clinical trials.

CRISIS INTERVENTION

Research is branching beyond anti-sickling agents to molecules that prevent sickle-cell interaction. Sickle cells do not act in isolation: not only do they stick to each other and obstruct blood vessels, but they also cause white blood cells, platelets and cells on blood-vessel walls to stick to each other. The result — blocked blood

vessels and inflamed tissues² — worsens the effect of the disease, says Marilyn Telen, director of the Comprehensive Sickle Cell Center at Duke University in Durham, North Carolina. They cause a common complication, called vaso-occlusive crisis, that causes periodic episodes of pain so intense that they often require hospitalization (see page S8) and which, over time, can cause permanent organ damage.

“It’s mind-boggling that there is still no specific therapy for vaso-occlusion,” says Paul Frenette, director of the Gottesman Institute for Stem Cell and Regenerative Medicine Research at Albert Einstein College of Medicine in New York City. On the basis of the number of drugs in the pipeline, he is hopeful that this will soon change.

One such drug, called MST-188, is thought to work by binding to the damaged membranes of red blood cells to make them less likely to stick to other cells. “It’s sort of like grease for your circulatory system,” says Brian Culley, chief executive of Mast Therapeutics, the biotechnology company in San Diego, California, that owns the drug. Like Aes-103, MST-188 had been shelved by its creators. In 2010, Mast Therapeutics acquired the company that owned MST-188 to reboot the drug’s development. Mast is now recruiting patients at medical centres around the world for a phase III trial, in which it aims to test MST-188 as a treatment for vaso-occlusion.

Other researchers are attempting to prevent blood-vessel blockages by targeting specific molecular interactions between cells. One of the most promising targets is a class of cell-adhesion proteins called selectins that, when activated on the surface of epithelial cells, cause white blood cells to stick to blood-vessel walls and have been implicated in the pathophysiology of sickle-cell disease in animals. The furthest along of these studies is being pursued by GlycoMimetics of Gaithersburg, Maryland, which has completed a phase II study of a drug called rivipansel that inhibits selectin molecules. In patients hospitalized for vaso-occlusion, rivipansel reduced the duration of the crises, shortened hospital stays, and decreased the amount of opioid pain medication by 83% compared with placebo. The company is now collaborating with Pfizer on a phase III trial.

Because vaso-occlusion involves a variety of processes, including pain and inflammation, combination therapy might be the best way to reduce its effects. But, Telen says, “we are hugely far away from that”, and notes that developing such a cocktail is daunting for a rare disease with little funding.

FETAL INSPIRATION

The only therapy currently approved to treat sickle-cell disease, hydroxycarbamide, acts in a different way from the drugs being developed. Rather than treating the acute pain of vaso-occlusive crises, hydroxycarbamide prevents long-term consequences of the disease primarily by stimulating production of the fetal form

of haemoglobin. This molecule is the dominant form at birth but is present only in very small amounts by the time a child is one year old. It functions like adult haemoglobin but, because it is produced by a different gene, is unaffected by the sickle-cell mutation.

“Our goal is to reactivate the fetal haemoglobin gene.”

So researchers are investigating other drugs that stimulate fetal haemoglobin. Some are new agents and others have already been FDA-approved for treatment of unrelated conditions. These include pomalidomide, a drug used against multiple myeloma, and decitabine, a drug used for a type of bone-marrow cancer. In each case, says Pace, “our goal is to reactivate the fetal haemoglobin gene, or to keep it active, if you treat a very young child”.

Beyond finding potential drug treatments, a major hurdle in getting promising therapies tested and approved is patient recruitment. Several clinical trials have stopped altogether because of low enrolment. “We’re still trying to figure out different ways to overcome the problem,” says Carlton Haywood, a bioethicist specializing in sickle-cell disease at Johns Hopkins University in Baltimore, Maryland. Studies suggest that the reasons why so few patients with sickle cell in the United States participate in clinical trials include a lack of trust in the medical establishment, either because of a history of discrimination or because of poor experiences in their own medical care, and parental unwillingness to expose children to an experimental drug.

But Haywood, who has sickle-cell disease, adds that it is also important for researchers to examine institutional factors that prevent enrolment, such as study designs that exclude too many people because of medical complications or if they have other diseases. Also, because the disease already causes people to take time off work or school, “factors like providing child care or transportation could make research participation less of a burden”. And he has found that people with sickle cell in the United States, most of whom are African-American, also report racial or ethnic discrimination in the health-care system, and it is primarily US patients who are being recruited to sickle-cell drug trials.

Haywood believes that researchers should focus on helping patients feel more engaged in clinical research. “We should feel like we’re an important part of the process”, he says — something that is particularly important when the pool of potential subjects is already so small. Ultimately, even the most promising treatments will not advance without patients willing to test them. ■

Courtney Humphries is a freelance science writer in Boston, Massachusetts.

1. Abdulmalki, O. *et al. Br. J. Haematol.* **128**, 552–561 (2005).
2. Manwani, D. & Frenette, P. S. *Blood* **122**, 3892–3898 (2013)



Q&A Graham Serjeant

Healthy progress

Physician Graham Serjeant reflects on how Jamaica's inhabitants have helped to advance our understanding of sickle-cell disease.

How have you been involved in sickle-cell disease research?

I moved to Jamaica 48 years ago, initially to develop services for patients with sickle-cell disease and to run the Medical Research Council Laboratories (Sickle Cell Unit) at the University of the West Indies in Kingston. I retired from that post about 15 years ago and have been working with a local charity called Sickle Cell Trust (Jamaica), where one of our most important projects — the Manchester project — focuses on prevention. Besides Jamaica, I have also worked on sickle-cell disease in Brazil, Saudi Arabia, India and a number of countries in Africa.

What does the Manchester project hope to achieve?

The long-term control and management of sickle-cell disease has to come in the form of prevention. Everyone has the right to knowledge of their haemoglobin genotype and to its significance (see page S2). They would then be empowered to make informed decisions.

The Manchester project seeks to inform a group of intelligent and motivated senior-school students of their haemoglobin genotype, then observe whether that knowledge influences their reproductive decisions and results in a significant reduction in

the number of babies born with sickle-cell disease. With support from the Jamaican ministries of health and education, we offered free, voluntary testing for haemoglobin genotype to students mostly aged 15–19 over a period of six years. We screened more than 16,000 students, almost 15% of whom carried abnormal genes. Every student received a laminated card, like a credit card, with their personal information and their genotype. Students with an abnormal genotype received additional information and counselling.

The next phase involved setting up newborn screening for south-central and western Jamaica — areas where the students were most likely to have their babies. The screening covers approximately 15,000 babies a year and so far, around 1,700 babies have been born to our informed cohort. The project appeared to be progressing well, until two mothers gave birth to babies with sickle-cell anaemia (who inherited a copy of the sickle-cell variant of the haemoglobin gene from both parents) within a period of two weeks. So far, of 845 females with the sickle-cell trait, four babies have been born with sickle-cell anaemia in 163 first pregnancies, which is close to the expected outcome had there been no intervention.

Have similar interventions been attempted in the past?

A smaller intervention known as the Orchomenos project was conducted in Greece between 1966 and 1970, but the results were not very encouraging. Four couples, in each of which both partners had the sickle-cell trait, had married. Two couples married with full knowledge of their genotypes and the related risks, and the other two had hidden their genotypes from their partners. But we have to persevere. By preventing as many cases as possible, we can ease the burden of providing clinical care for the patients that continue to be born.

How big is the current burden of sickle-cell disease?

Diarrhoea, infections and malaria may cause more deaths than sickle-cell disease, but as we begin to control these and other conditions, sickle cell is going to emerge as an increasingly larger problem. Not only are the numbers daunting — a quarter of a million babies are born with sickle-cell anaemia every year in sub-Saharan Africa alone — but once you have the disease, the problems may be life-long and require good monitoring and clinical care.

Has the global public-health response been sufficient?

In the past five years, there has been growing recognition of the importance of the disease, and the realization that much can be done to prevent many of the associated health problems. Instead of sitting back helplessly and saying, "This is a genetic disease and we cannot

cure it", the truth is that you actually don't have to cure it. If people can survive to the age of 80, as some of our patients in Jamaica do, then that is a disease that we can live with. The National Sickle Cell Anemia Control Act, passed in 1972 in the United States, was a major stimulus for raising awareness of the need for education, counselling and population screening.

But lack of political will continues to significantly impede progress. Faced with the multitude of clinical problems, some health-care personnel in Africa may say: "So, the children die? Well, thank God for that!" From their perspective, sickle cell shrinks into the background compared to the more immediate and overwhelming

"Once you have sickle-cell disease, the problems may be life-long and require good monitoring and clinical care."

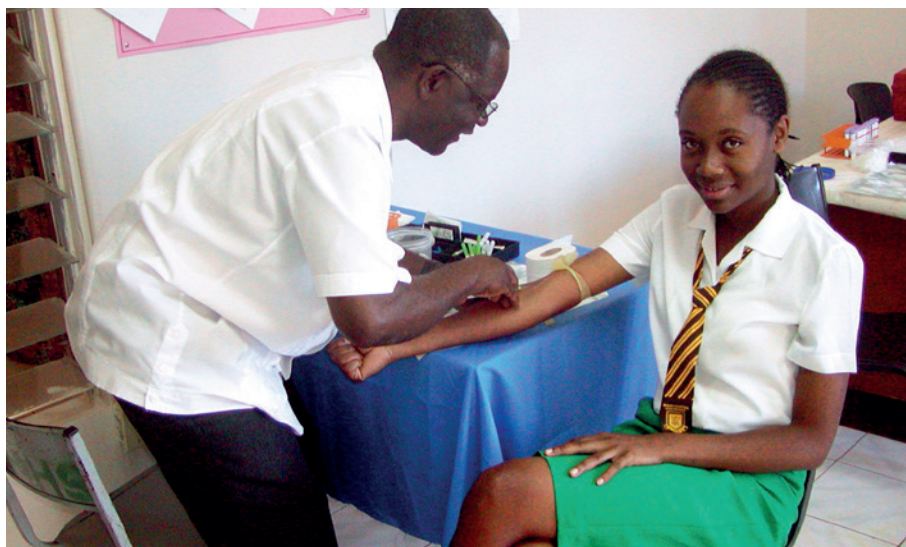
problems of HIV, malnutrition, infections, septicaemia and Ebola, among others. But in the long term, they are going to be left with sickle-cell disease, and services have to be developed. There have been some positive developments with the establishment of dedicated centres in Lagos, Nigeria and Kumasi in Ghana, but it is going to be a long haul.

Meanwhile, India has made great progress. In the past 15 years, three states have established state-managed sickle-cell control programmes. They are: Gujarat, Maharashtra and Chhattisgarh, where sickle-cell disease is endemic among the scheduled castes and scheduled tribes (historically marginalized and socioeconomically disadvantaged groups identified by the Indian Constitution for special consideration). I am not aware of any African country that has done that.

How have interventions in recent years improved patients' quality of life?

We can now diagnose the disease at birth; the technology is simple and the population is accessible. Major causes of morbidity and mortality — especially in the second six months of life — such as pneumococcal septicaemia, acute splenic sequestration, acute chest syndrome, and even stroke can be prevented by simple interventions.

The painful crises, which become a problem in adolescence, are more difficult to manage. But we have models of care that work well in our Jamaican population, which involve improving the capacity of patients to cope with the pain by talking them through it. This may not be possible in societies in which doctors have limited time or feel obliged to do more for their patients. In such situations, a shot of morphine or the preventative drug hydroxyurea may be the solution, but not for our Jamaican patients in whom our alternative methods have worked,



Screening for the sickle-cell gene as part of the Manchester project, run by Sickle Cell Trust (Jamaica).

especially given the relatively high cost of hydroxyurea and concerns about the potential side effects. A crucial element of the success in Jamaica has been the provision of diagnostic services in dedicated sickle-cell centres that are staffed with experts who are familiar with the disease and their patients.

The median survival for sickle-cell anaemia is less than five years in sub-Saharan Africa (although formal survival estimates are not available), but in Jamaica it is 55 years. This is still about 20 years shorter than the general population, but our early childhood interventions, which we began implementing in the early 1980s, have yet to feed through to be reflected in overall survival rates.

What has been the most important thing that you have learned about the disease?

When I first came to Jamaica in 1966, our sickle-cell clinic at the University Hospital of the West Indies was attended by patients with sickle-cell anaemia who were between 30 and 40 years old — an observation at total variance with the traditional descriptions of the disease, which referred to it as almost always causing death in childhood. One possible hypothesis was that these traditional descriptions were based on a highly selected, symptomatic group of patients available to the doctors writing on sickle-cell disease. We tested this hypothesis with support from the Wellcome Trust in London by driving all over Jamaica in a mobile clinical unit to locate patients who had been known to the University Hospital but had not been seen for ten years. We found more than half of the patients, all of whom said that their symptoms had improved with age, so they had no need to waste their time seeing a doctor at the hospital.

That was a big surprise, and we realized that once you can get out into the community, you get to see the other side of the disease. It was a supreme example of the value of Jamaica

as a clinical laboratory, being an island with a strong extended-family system and many patients, 5,500 of whom were under regular observation at the sickle-cell clinic.

These early observations led to the Jamaican Cohort Study in 1973, when the Sickle Cell Unit began screening 100,000 consecutive, non-operative births at Victoria Jubilee Hospital in Kingston, and detected 550 children with different genotypes of sickle cell who are being followed for life. The cohort study has helped to define the true natural history of sickle-cell disease and develop cost-effective interventions to improve the outcome of the disease.

Can the wide variation in disease severity be explained by genetics?

One of the classic methods for distinguishing genetic from environmental factors in disease is studies of identical twins. We identified nine pairs of identical twins in Jamaica, of whom six had sickle-cell anaemia. Analysis showed that identical twins were similar in terms of haematology and growth, but discordant in virtually all other clinical features. Even though these twins were genetically identical, their experience of the disease was very different. Genetics accounted for only a small portion of the variability in the disease.

Interestingly, that was not a message that people wanted to hear in the post-genome era, and we faced considerable difficulty in getting those results published. But to me, the results were hopeful because we cannot change someone's genes but we can change his or her environment. There is also a huge psychological component to consider, such as the role of stress in inducing serious complications. I have seen people die in a painful crisis during their university exams. The tremendous clinical variability of the disease still remains one of the biggest unknowns. ■

INTERVIEW BY SMRITI MALLAPATY



ANDRÉ DALOBA

NEUROBIOLOGY

Life beyond the pain

People with sickle-cell disease experience bursts of debilitating agony from birth. But researchers have made a promising discovery that could ease these excruciating episodes.

BY BIANCA NOGRADY

At first glance, a child with sickle-cell disease may appear healthy. But what you cannot see — and what dominates the child's life — are bursts of searing, crippling pain that strike without warning. These episodes require trips to hospital, and even the highest doses of the most powerful opioid drugs typically provide only temporary, incomplete relief¹.

As children with sickle-cell disease grow older, the pain worsens. They continue to have intense and debilitating acute attacks, and by the time they reach adulthood nearly one-third of them will also experience steady, unrelenting, background pain every day of their lives.

Sickle-cell pain has been thought of as a distress cry of oxygen-starved tissues, as the blood vessels supplying them become blocked and damaged by distorted blood cells. But in the past five years, a proliferation of pre-clinical and early clinical research is forcing a rethink of this assumption. Each new discovery is revealing a jigsaw puzzle in which damaged blood vessels, oxygen-deprived tissues, inflammation, opioid tolerance and hyper-sensitivity are intrinsically connected. Together, these processes present a much more complex

picture of sickle-cell pain than researchers had previously believed, one of a damaged and persistently activated nervous system.

A SURPRISE DISCOVERY

In 2013, pain specialist Diana Wilkie, who works at the College of Nursing at the University of Illinois in Chicago, and her colleagues enrolled 18 people with sickle-cell disease in a phase I trial of an agent that they hoped might offer some relief². One of those participants was a woman with sickle cell who had been hospitalised with severe pain 38 times the previous year and could not get further prescriptions of pain-relieving, or analgesic, opioid medication because of limits put in place by her medical insurance. The trial's initial findings were positive: "She had profound analgesia with the study drug", as did nearly half of the study participants, says Wilkie. "She was so happy."

The drug is not a new opioid, or a novel painkiller. It is a widely used antipsychotic medication called trifluoperazine. Why would medicine used to treat the symptoms of schizophrenia have an effect on pain caused by a blood disease? The answer to that question lies in a mystery that Wilkie and her colleagues have been trying to unravel for some time.

The first indication that a non-opioid might mitigate sickle-cell pain came from a study³ that Wilkie's team published in 2010, in which they asked 145 people with sickle-cell disease to describe their pain using a standard form called the McGill Pain Questionnaire. The researchers expected participants to characterize their pain using terms generally associated with tissue damage caused by oxygen starvation — words such as 'pounding'. But to their surprise, more than 90% also picked descriptors commonly used to describe pain that results from damaged or diseased nerves — words such as 'aching', 'shooting', and 'stabbing'. "We realized that there was a real phenomenon here, and this realization came from the patients' language of their own pain," says Wilkie.

Other evidence was emerging to undermine the conventional belief that sickle-cell pain stemmed from 'vaso-occlusive crises' — the name given to episodes during which blood vessels become blocked by malformed (or 'sickled') red blood cells. Deepika Darbari, a paediatric haematologist at the Children's National Medical Center in Washington DC, says that even people with sickle cell who receive regular blood transfusions to reduce the risk of blood clotting and stroke — and therefore have extremely low levels of sickling — still have pain. "If the

KALPNA GUPTA vaso-occlusive crisis was the only thing causing pain, then these patients should not have any pain because you have taken the sickle cell out of the equation,” Darbari says. “So, what else is causing it?”

FEELING SENSITIVE

Mouse studies provide a glimpse into the nervous-system pathology that may underlie some of this long-term pain. Mice cannot communicate pain as directly as a human might, but researchers can still measure an animal's discomfort. They do this by measuring the mouse's grip strength, which is weaker when it is in pain, or by monitoring how quickly it draws its paw away from a stimulus.

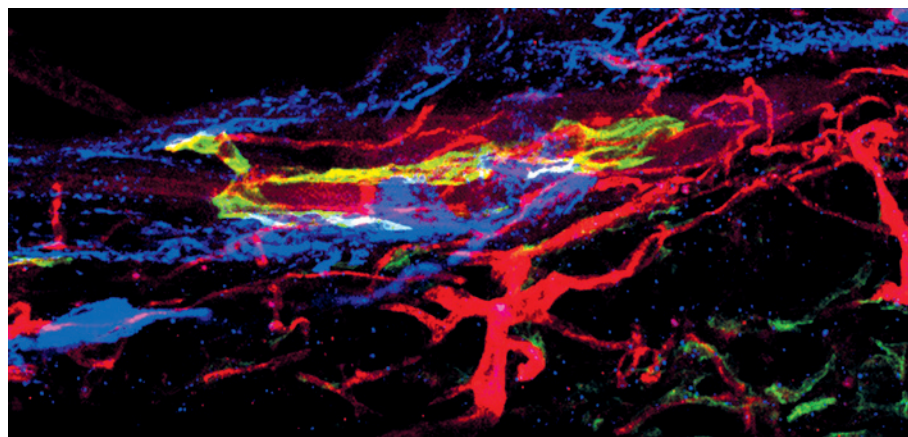
To understand the mechanisms of pain in sickle-cell disease, scientists such as Kalpna Gupta, a haematological oncologist at the University of Minnesota in Minneapolis, are examining every level of the pain system in sickle-cell mice and control mice. These studies investigate the entire biological mechanism, from the peripheral nerve fibres in the skin through to the brain and central nervous system, where pain is perceived.

Gupta's work has revealed key differences between the control mice and sickle-cell mice. Not only does the skin of the sickle-cell animals contain fewer nerve fibres, but the nerve fibres that do exist have structures that are arranged in a haphazard and distorted way. The nerves of sickle-cell mice also react much more strongly to two proteins known to cause neural inflammation and increased sensitivity to pain. To Gupta and her colleagues, these observations all point to peripheral nerve damage, possibly the result of inflammation, which in turn leads to sensitization of the nervous system and chronic pain.

Gupta's team has also found⁴ increased production of several other substances — the inflammatory cell-signalling molecule interleukin-6, the enzyme COX-2 and a protein of the immune system called toll-like receptor 4 — in the animals' spinal-cord nerves, which they propose would increase nerve inflammation and sensitisation.

Another of Gupta's findings may help to explain why the effects of opioid drugs fade so quickly for people with sickle cell, and why they need higher doses than people with other conditions who have similar levels of pain. Nerves in the skin and spinal cords of the sickle-cell mice showed a reduced level of the μ -opioid receptor, which enables opioid painkillers such as morphine to take effect. According to Gupta, sickle-cell mice require greater amounts of opioids for pain relief than are normally used in mouse models of pain.

Opioid tolerance in people with sickle-cell disease has been blamed on the fact that the extreme discomfort associated with the condition requires huge doses of opioid medication, leading to an assumption that people become resistant to the effects of the drugs and enter



Abnormal clusters of blood vessels (red), nerves (blue) and lymph (green) in the skin of sickle-cell mice.

a vicious cycle of escalating doses and continued pain. Yet Gupta's sickle-cell mice had no previous morphine exposure and could not have developed tolerance. She concluded that in mice, opioid tolerance may be another element of sickle-cell biology. Gupta says the findings help to explain the problem of opioid use among people with sickle-cell disease, which is viewed not only as a scientific issue but also as a social one — fear of encouraging addiction may mean that some doctors do not prescribe high doses of opioids to their patients.

Further evidence that pain is related to nerve damage came from the discovery that the sickle-cell mice were much more sensitive to cold, heat and touch than control mice. Amanda Brandow, a paediatric haematologist and oncologist at the Medical College of Wisconsin in Milwaukee, has found the same sensitivities to heat and cold — but not touch — in humans with sickle-cell disease⁵. For example, her study participants experienced discomfort at temperatures much closer to the neutral baseline of 32 °C than healthy, race-matched controls.

NEW OPPORTUNITIES

One possible explanation for this extraordinary temperature sensitivity is a receptor called TRPV1 (transient receptor potential vanilloid), which is found on certain sensory neurons. When Brandow and her colleagues blocked the TRPV1 receptor in sickle-cell mice, they found that it reversed the animals' sensitivity to touch. The TRPV1 receptor helps to activate an enzyme (CaMKII α) that is emerging as a major player in sickle-cell pain.

CaMKII α has been shown to interact with certain nerve receptors called NMDA receptors. This attracted the attention of Jim Wang, a pharmacologist at the University of Illinois at Chicago, and a collaborator of Wilkie's, because NMDAs have a key role in the development of long-term pain.

This, combined with the fact that CaMKII α is expressed in clusters of nerve cells in the spinal cord that process pain signals, suggests that CaMKII α may be a new target for pain relief in sickle cell and some other conditions.

Instead of creating a new CaMKII α inhibitor from scratch, Wang wanted to see whether other drugs that are currently available might be suitable. “We started, on a computational model, just looking through the clinically approved drugs that could modulate this enzyme activity,” he says.

Wang's systematic search led him to trifluoperazine. His phase I trial of this drug, in partnership with Wilkie and others, halved the pain levels in 8 of the 18 study participants, without any need for extra opioid pain relief. Moreover, Wang says, trifluoperazine may have the added effect of reducing people's opioid tolerance levels so they are once again responsive to those medications. The fact that trifluoperazine worked for some people but not others may have been a result of genetic differences. Or, Wilkie says, it may have been because different patients received different doses, some of which may have been too low to cause a noticeable effect.

The discovery that sickle-cell pain involves neural pathways opens a new chapter in the search for treatments. After decades of opiate-based therapies, Brandow says, the past few years have seen some significant contributions to understanding sickle-cell pain. “The more we understand about the neurobiology, the more people we can get interested in studying it who have experience in pain research,” she says. “From that may come novel therapies, which we desperately need.” ■

Bianca Nogrady is a freelance science writer in Sydney, Australia.

1. Ballas, S. K. *Blood* **116**, 311–12 (2010).
2. Molokie, R. E. et al. *Eur. J. Pharmacol.* **723**, 419–424 (2014).
3. Wilkie, D. J. et al. *J. Natl. Med. Assoc.* **102**, 18–27 (2010).
4. Kohli, D. R. et al. *Blood* **116**, 456–465 (2010).
5. Brandow, A. M. et al. *Am. J. Hematol.* **88**, 37–43 (2013).

PERSPECTIVE



We need a global solution

The disparity in outcomes for children with sickle-cell disease in developed and developing countries is an injustice, says **Isaac Odame**.

It has been more than a century since sickle-cell disease was first described in the medical literature, and half a century since its genetic basis was understood at the molecular level. Medical management remains largely suboptimal, however, particularly in low-income countries where the burden of disease is heaviest.

From its origins in sub-Saharan Africa, the Arabian peninsula and the Indian subcontinent, the sickle-cell variant of the haemoglobin gene has spread through the rest of the world. Using geostatistical models, epidemiologists estimate that more than 300,000 babies with two copies of the sickle haemoglobin gene are born every year¹. This is probably an underestimate, but it provides the best current approximation of populations afflicted with the disease. More than three-quarters of those affected are born in Africa, with half the global burden shouldered by just three countries: Nigeria, India and the Democratic Republic of Congo.

In high-income countries, which house less than 10% of the global burden, early mortality has been significantly reduced over the past few decades, with more than 90% of people with sickle cell surviving beyond 20 years of age. This can be attributed to widespread implementation of a range of interventions, including newborn screening, penicillin prophylaxis (because sickle cell increases susceptibility to pneumococcal infection), pneumococcal vaccination and parental education. By contrast, the World Health Organization (WHO) estimates mortality of those younger than five to be more than 50% in low-income countries, and has estimated that sickle cell accounts for up to 9% of all deaths in children under five in sub-Saharan Africa.

Causes of mortality in these patients include malaria-associated severe anaemia, pneumococcal and other bacterial infections. In addition, the disease is associated with significant health challenges including recurrent debilitating pain, chronic anaemia and stroke.

LIMITED RESOURCES

In low-income countries, limited resources for diagnosis and treatment, combined with a dearth of government strategies to combat the disease, have led to a paucity of care. This is only compounded by the lack of public understanding of sickle-cell disease, which perpetuates social stigma and myths about disease causation and results in few people seeking appropriate treatment. In 2010, the WHO Regional Office for Africa commissioned a strategy document that spelled out guidelines for actionable steps to combat sickle-cell disease, but these are yet to be translated into government action.

Part of the United Nations' Millennium Development Goals — a decade-long initiative to help the world's poorest citizens — aims to reduce mortality rates among children under five by two-thirds by 2015. Although most countries in Africa have yet to achieve this objective, the initiative has resulted in visible improvements in areas such as nutrition, immunization, management of diarrhoea, malaria control

and antibiotic coverage — interventions that are reducing mortality in children under five and causing an epidemiologic transition whereby some children with sickle-cell disease who would have died undiagnosed in early life now survive and need continuing care². So the burden of sickle-cell disease will continue to rise, placing an increasing strain on limited health-care resources. We must find affordable, evidence-based solutions that can be integrated into existing health-care systems to ensure their sustainability.

To address sickle-cell disease in low-income countries requires a number of feasible steps. First, the diagnostic barrier needs to be broken. Current laboratory methodologies are too costly to enable equitable and timely diagnosis. Low-cost, rapid, point-of-care diagnostic tools, which are currently being developed, are needed to facilitate early detection. Second, simple, life-saving interventions should be

integrated into primary-care delivery systems³. The Gavi vaccine alliance, a public-private global partnership committed to improving access to immunization, has embarked on an initiative to expand primary pneumococcal vaccination that will be beneficial to children with sickle-cell disease in poor countries. Widespread introduction of hydroxyurea — a relatively affordable, disease-modifying drug — would help to improve quality of life and prolong survival. Clinical trials to determine the safety, dosing and effectiveness of hydroxyurea therapy in low-income countries have begun in Africa and are scheduled to end in 2016. If successful, the implementation and distribution could be scaled up using the same methods that proved effective for improving access to antiretroviral therapy for HIV-positive

individuals. Research has already shown which interventions are effective. Now, emphasis must shift to implementation so that we can determine how to make them work for every child and adult.

We need public-awareness campaigns to combat stigma and misunderstandings about sickle-cell disease, and to provide opportunities for genetic counselling and initiatives that promote disease avoidance. Many effective interventions are affordable and within grasp if there is a will to act, and civil society and advocacy organizations in these low-income countries should be emboldened to lobby their governments to place sickle-cell disease high on their agenda. Finally, broad public-private partnerships are critical to finding and implementing sustainable solutions that reduce the global burden of the disease. Only then can we hope to bring an end to this century-long injustice. ■

Isaac Odame is medical director of the Global Sickle Cell Disease Network and professor of Paediatrics at the Hospital for Sick Children and University of Toronto, Canada.
e-mail isaac.odame@sickkids.ca

1. Piel, F.B. *et al.* *The Lancet* **381**, 142–151 (2013).

2. Weatherall, D. *et al.* *Blood* **105**, 921–923 (2005).

3. Aygun, B. & Odame, I. *Pediatr. Blood Cancer* **59**, 386–390 (2012).



DNA sequences showing the sickle-cell disease mutation (marked with an asterisk, top) and the sequence corrected (below) using gene-editing technology.

GENE THERAPY

Editorial control

Correcting the genetic error in sickle-cell disease might be as simple as amending text.

BY KATHARINE GAMMON

Tiny changes in DNA can have huge consequences. For years, scientists have been trying to 'fix' these mutations in the hope of treating and potentially curing some of humanity's most devastating genetic diseases. After some tragic early setbacks (see *Nature* **420**, 116–118; 2002), techniques that allow precise genetic manipulation have created a surge of research.

Although most existing treatments for genetic diseases typically only target symptoms, genetic manipulation or 'gene therapy' goes after the cause itself. The approach involves either inserting a functional gene into DNA or editing a faulty one that is already there, so the conditions most likely to prove curable are those caused by a single mutation. Sickle-cell

disease is a perfect candidate: it is caused by a change in just one amino acid at a specific site in the β -globin gene. This results in the production of abnormal haemoglobin proteins that cause the red blood cells that house them to twist and become sickle shaped. The distorted cells get sticky, adhere to each other and block blood vessels, preventing oxygenated blood from flowing through.

Gene therapy has been used successfully in a handful of patients with immune disorders, and sickle-cell disease is among researchers' next targets. The most advanced of these projects is slated to begin clinical trials by the end of the year, and other trials are set to follow. The approaches being developed to treat sickle-cell disease take one of two forms. Conventional gene therapy, also known as gene addition, typically involves inserting new genes. Usually, a harmless virus

is modified with the gene to be inserted, and this 'viral vector' is mixed with cells from the patient *in vitro*. The virus searches out the cells and inserts the gene into the cells' DNA, after which the cells are transplanted into the patient. Conversely, gene editing is more nuanced: in a molecular cut-and-paste, researchers cut out the faulty DNA sequence and then insert a piece of laboratory-created DNA. In both approaches, the modified DNA dictates the formation of a normal, working protein.

In sickle-cell disease, the only cells that need their DNA edited are blood stem cells — also known as haematopoietic stem cells — which are found in bone marrow. These cells continually form new red blood cells to replace those that are lost, and reprogramming just a small fraction of them will create enough perfectly formed red blood cells to eliminate disease

MOLECULAR CUT-AND-PASTE

Three different gene-editing techniques could allow researchers to fix the single mutation in the haemoglobin gene that causes sickle-cell disease.

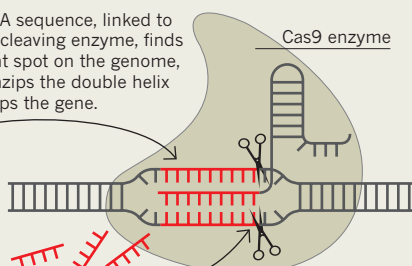
CRISPR CUTS

Clustered regularly interspaced short palindromic repeats (CRISPR) use RNA to guide precise cuts in the genome.

- 1 A section of RNA is engineered to target a specific region of DNA.

Guide sequence

- 2 The RNA sequence, linked to a Cas9 cleaving enzyme, finds the right spot on the genome, then unzips the double helix and snips the gene.

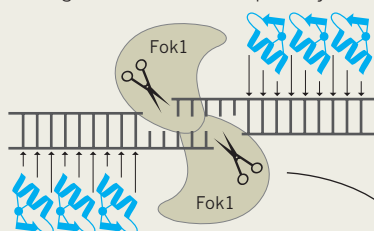


- 3 Before the DNA is repaired, small amounts of engineered DNA can be inserted to alter the gene.

ZFNs

Zinc finger nucleases (ZFNs) are used in pairs to cut either side of the double-stranded DNA.

- 1 Zinc fingers (blue) recognize specific DNA sequences — typically three base-pairs per finger — and must match precisely.



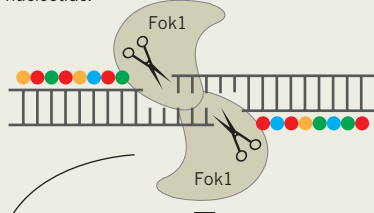
- 2 The fingers bind to matching DNA. The Fok1 enzymes bind to each other and snip out the offending gene.

- 3 As the cells repair themselves, they incorporate a healthy version of the gene (red).

TALENs

Transcription activator-like effector nucleases (TALENs) work like ZFNs. Using one long matching sequence allows gene recognition to be precise.

- 1 TALENs are also used in pairs, one for each DNA strand. They consist of amino-acid sequences (coloured dots), each of which binds to a single nucleotide.



- 2 When the long sequences match up on either side, the Fok1 enzymes bind to each other and snip out the unwanted gene.

symptoms. “Achieving genome editing via direct repair of blood stem cells represents a high hurdle,” says George Daley, director of the Stem Cell Transplantation Program at Boston Children’s Hospital in Massachusetts, “but perhaps not an impossible one.”

Although these approaches are promising, several important issues must be addressed before they can be used to treat patients, such as ensuring that the therapies accurately hit their targets and do not cause irreparable harm to the cells or introduce additional genetic information that could cause problems such as cancer.

INJECTING GENES

Gene addition is poised to become the first sickle-cell gene therapy to be tested in humans. At the regenerative medicine and stem cell research centre of the University of California, Los Angeles, molecular geneticist and physician Donald Kohn is developing protocols for a clinical trial of this technique that is due to start enrolling patients by the end of 2014. Doctors will first harvest bone marrow from the hip bones of patients with sickle-cell disease and then extract haematopoietic stem cells from the marrow. Using a viral vector,

they will insert a new, working haemoglobin gene into the cells’ DNA; the old, faulty haemoglobin gene will still be present, but it will go silent as the new gene takes over. The modified cells will then be infused back into the patient’s bloodstream and will migrate to the bone marrow, where they can provide a continual source of healthy red blood cells.

Kohn says that this approach has the potential to cure sickle-cell disease, and with significantly fewer side effects than a bone marrow transplant — currently the only cure (see page S14). He has tested the technique by injecting modified human haematopoietic stem cells into mice, and found that they were free of sickle cells 2 to 3 months later¹. The limiting factor in mice, Kohn says, is that they can only sustain human grafts for that long. In humans, he thinks the correction should last a lifetime — as long as 50 to 70 years.

One of the challenges in treating sickle-cell disease with gene therapy is that it is necessary to extract bone marrow to retrieve haematopoietic stem cells. With most other diseases, patients can be given drugs that entice these cells to leave the marrow and enter the bloodstream, where they can be easily harvested. But

in patients with sickle-cell disease, these drugs can trigger sickle-cell crisis, an acutely painful episode during which the damaged cells stick together and block blood vessels; the crisis can be accompanied by anaemia, chest pain, difficulty breathing, blood trapped in the spleen and liver, even stroke. So researchers must harvest the bone marrow itself, which can be difficult and slow, and limits the number of cells that can be collected at one time. Kohn says that they still do not know whether this approach will yield enough haematopoietic stem cells for reprogramming. And, like other bone marrow transplant procedures, the patient still needs to undergo chemotherapy to kill off the remaining bone-marrow cells to help the genetically altered ones survive once they are reintroduced into the body.

TALENTED FINGERS

Further away from clinical trials, but potentially a lot more exciting, is gene editing. The concept was introduced in the 1990s, when artificial DNA-cutting enzymes known as zinc finger nucleases (ZFNs) were first engineered. ZFNs bind to a specific section of DNA and create a break at both ends (see ‘Molecular cut-and-paste’). Cells will start to repair the break, at which point a specific sequence of laboratory-made DNA can be slotted into the gap. After the DNA is repaired, the cells start to create healthy copies of the gene.

In parallel to his work on gene addition, Kohn is exploring the use of ZFNs to edit sickle-cell genes. In collaboration with the firm Sangamo BioSciences in Richmond, California, he has shown that around 7% of haematopoietic cells can be repaired in culture using this technique, using a viral vector to get the ZFNs into the cells. Because the repaired cells continue to replicate, this small proportion could be enough to eventually produce a sufficient amount of working red blood cells. Kohn says that patients have shown major improvements when just 10–20% of their donor cells successfully engrafted and started to make new, healthy cells.

The advantage of gene editing over gene addition (a less complex approach) is that it provides an actual fix rather than a work around. But ZFNs are expensive and difficult to program. In 2010, a gene-editing protein called TALEN (transcription activator-like effector nuclease) was developed, which uses a similar mechanism as ZFNs but is cheaper and easier to work with. It was quickly adopted for use in sickle-cell disease.

At the Salk Institute for Biological Studies, in La Jolla, California, stem-cell biologist Juan Carlos Izpisua Belmonte uses TALENs in concert with viral vectors called HDAdVs (helper-dependent adenoviral vectors) to correct the sickle-cell mutation. Instead of harvesting haematopoietic stem cells from bone marrow, Izpisua Belmonte’s team takes easily harvestable cells, such as blood, skin or fat

cells, and then turns them into induced pluripotent stem (iPS) cells, which can be converted into any cell type. The researchers correct the haemoglobin gene defect *in vitro* using gene editing, then differentiate the repaired iPS cells into blood stem cells. From there, the researchers have a couple of choices. The repaired cells could simply be infused into a patient's bloodstream, where they would make their way into the bone marrow and start to make healthy haematopoietic cells.

But Izpisua Belmonte is also working on a cure that could work inside the bone marrow itself. His team is combining TALENs with a different viral vector, HDAdVs, to boost the success rate of gene editing, and the researchers are working on a plan to administer their hybrid vector directly into the bone marrow, so the genetic fix would take place inside the patient's body. Although each infusion into the marrow might correct only 1% of the cells, ten such procedures over the course of several months — something Izpisua Belmonte and his research associate Mo Li think is feasible in terms of time and cost — could alleviate the symptoms of sickle-cell disease. "Little by little, you are correcting the disease *in vivo*," says Izpisua Belmonte. So far, this 'hybrid vector' technique has shown promising efficacy in umbilical-cord blood stem cells.

Sickle-cell disease results when both copies of the haemoglobin gene are faulty, and fixing just one of the genes is sufficient to make a big health improvement. As Li points out, people who carry one copy of the mutated gene, a genetic condition referred to as 'sickle-cell trait', do not show symptoms. "In fact, many of the world's best sprinters have the sickle-cell trait, he says. "Our approaches will most likely restore one mutated copy to its wild-type sequence, leaving the other copy untouched."

CRISPRs (clustered regularly interspaced short palindromic repeats) are the most recent addition to the gene-editing toolbox. Whereas ZFNs and TALENs use a protein to lock on to a specific section of DNA, CRISPRs use a 'guide RNA'. These guide RNAs are much easier to program than the proteins in TALENs and ZFNs, as well as being cheaper and more efficient. CRISPRs also make it possible to perform multiple genetic manipulations in one go. CRISPRs work in combination with the Cas9 (CRISPR-associated 9) nuclease: after the CRISPR locks on to the target gene, Cas9 snips both strands of the DNA, disabling the gene. The approach is less than two years old, yet many researchers are now working with CRISPRs in parallel with other *in vitro* techniques.

There are safety hurdles to be overcome before gene editing is used in humans, especially because it involves a permanent change in the genome. The thorniest issue is 'off-target activity' — unintended changes to the genome away from the target gene.



A researcher corrects a mutation in the β -globin gene that causes sickle-cell disease.

Gang Bao, a biomedical engineer at the Georgia Institute of Technology in Atlanta, is developing gene-editing strategies for sickle-cell disease and is paying particular attention to the challenge of limiting off-target effects. He notes that if erroneous cuts happen in a cancer-causing gene, they could potentially trigger tumour growth. Even a rate of off-target activity lower than 1% could still pose serious health risks. So that the technology can move forward, researchers need to have a better understanding of off-target effects. There are two main issues: determining exactly where the off-target cuts occur and at what rate.

Bao's group has created software to predict where the off-target effects might occur for the different gene-editing techniques. In a paper published in May, his team reported that their software predicted 114 potential off-target sites across the whole genome for the CRISPR/Cas9 system, and experiments confirmed 15 of them by sequencing the cleaved DNA².

Izpisua Belmonte's team is also looking at the rate of unwanted mutations caused by gene-editing techniques. The group created iPS cell lines and then edited half of the cells using HDAdVs and TALENs³, but left the other half unedited. The edited cells had no more mutations than the unedited ones, indicating that — in contrast to Bao's findings for CRISPRs — the use of TALENs does not seem to make cells any less safe. Although human testing is still a few years off, they say that these results give them optimism about the potential for gene editing to work.

The other major challenge for gene-therapy researchers is ensuring that the edited stem cells survive and generate healthy red blood cells after they are reinserted into the bone marrow. Edited cells often die because of the amount of stress they undergo during therapy. Researchers might be able to improve the cell-survival rate by delivering other types of cells at the same time, and the speed of gene editing

also seems to be important: the longer the cells are cultured *in vitro*, the less likely they are to survive. "Let's see if we can perform the whole procedure in four hours instead of four days," says Bao.

FUTURE REPAIR TOOLBOXES

Based on his research so far, Bao thinks that CRISPRs are the best method for generating DNA breaks, but they are also more likely to cause off-target activity. TALENs are less efficient than CRISPRs, but they seem to have fewer off-target effects. The rate of on-target activity for CRISPRs is between 40% and 80%, whereas the on-target rate for TALENs is between 20% and 50%, Bao says. The rate of off-target activity varies depending on the type of cells and the nuclease used. "If we have a way to overcome the off-target and [cell-survival] problems, CRISPR is a very promising technology," he says.

Kohn has compared ZFNs, TALENs and CRISPRs, and thinks all three have therapeutic potential for patients with sickle-cell disease. The techniques are all good at slicing DNA; now the remaining challenges are delivering them to the target cell and accurately repairing the gene after the break.

Ultimately, for sickle-cell gene therapy to become reality, the details must be sorted out on a large scale. Tinkering with human genes can yield both devastating and remarkable results, and the difference between the two often lies in a single nucleic acid of a single gene. This places a heavy responsibility on the shoulders of every researcher in the field, but the vast potential of gene therapy makes that burden worthwhile. ■

Katharine Gammon is a freelance writer in Santa Monica, California.

1. Romero, Z. *et al.* *J. Clin. Invest.* **123**, 3317–3330 (2013).
2. Fine, E. J. *et al.* *Nucleic Acids Res.* **42**, e42 (2014).
3. Suzuki, K. *et al.* *Cell Stem Cell* **15**, 31–36 (2014).



ROBERT HARRIS/ST GEORGE'S HEALTHCARE NHS TRUST

A woman donates blood stem cells at St George's Hospital in London. Stem cells harvested here can help patients with sickle-cell disease and leukaemia.

STEM CELLS

Creating a cure-all

Stem-cell transplantation can cure sickle-cell disease, but so far this has been limited to a lucky few. That is changing fast.

BY ANDREW R. SCOTT

Like all patients with sickle-cell disease, Stephanie Alvarado-Ross had to cope with excruciating pain, anxiety and stress, and faced the likelihood of an early death. But after 20 years of suffering, that all changed in November 2013 when stem cells from her brother's bone marrow were infused into her bloodstream at the Dana-Farber/Boston Children's Cancer and Blood Disorders Center in Boston, Massachusetts. The donated cells repopulated Stephanie's bone marrow and produced healthy blood cells, delivering what will hopefully be a permanent cure. "I am only now beginning to realize what difference the transplant has made in my life," says Alvarado-Ross, who is now free of the disease. "I will no longer have to manage or worry about increasing complications or crises due to a chronic condition."

Alvarado-Ross received her life-changing transplant from a team led by Leslie Lehmann, clinical director of the centre's stem-cell transplantation programme. So far, Lehmann says, stem-cell transplant therapy, in which a patient's diseased bone marrow is

killed off before being re-established with stem cells from a healthy donor's bone marrow, has cured about 600 people worldwide. But the procedure is not available to everyone¹. It has generally been offered only to children and adolescents, because the disease causes organ damage that puts older patients at greater risk of life-threatening complications, including pulmonary hypertension, kidney failure and neurovascular problems.

Alvarado-Ross just squeaked in: at 20, she was nearly considered too old, but she was healthier than most patients of her age. She was also lucky to have a fully immunologically matched sibling donor — someone with identical human leukocyte antigens (HLAs) on the surface of their cells. Full HLA matching has been an absolute requirement for attempting stem-cell transplantation to ensure that the donor and host cells are compatible, but fewer than 10% of people with sickle-cell disease have such a donor available.

Transplantation has also involved the drastic approach of first completely destroying the patient's bone marrow — and hence their immune system — with chemotherapy and radiation, providing a clean home for

the new cells. This step increases both the short- and long-term toxicity of the treatment, however, and if the donor cells fail to engraft, the patient's outlook is bleak. Other potential problems include graft-versus-host disease, in which the donated stem cells attack the recipient's cells, as well as infections and other complications. All this makes stem-cell transplantation risky, says Lehmann: there's a 90% cure rate but a 5% mortality rate.

But recent research is making the procedure more inclusive and safer. Small-scale trials of gentler techniques, which do not entirely destroy the recipient's bone marrow and use established drugs, are broadening the age range of potential recipients and increasing the pool of potential donors. It may not be long before transplants become a standard option for many more people, maybe even for fetuses in the womb.

GENTLE CONDITIONING

The first hint that a milder approach might work came from examining the blood of children whose bone marrow had been destroyed in the usual way before the transplant. In a small proportion of cases, donor cells were

found living alongside native cells that had managed to survive. Having as few as 11% of the donor cells in this 'mixed chimaera' state was sufficient to overcome the disease, and there was no evidence of graft-versus-host attack or immune rejection. The reasons for this mutual tolerance of host and donor cells remain unclear, says John Tisdale, a clinical researcher at the US National Heart, Lung, and Blood Institute in Bethesda, Maryland, and his team is investigating it. In the meantime, they have replicated the mixed-chimaera state in adults with a mild pretransplant regime that Tisdale likens to "killing weeds and replanting in their place, rather than killing off everything in a garden and starting all over again."

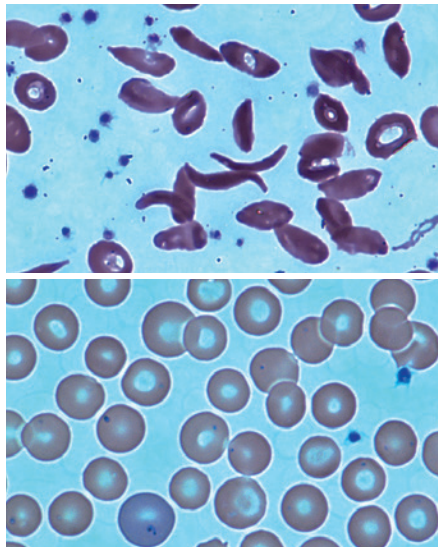
Tisdale's team replaced the combination of high-dose radiation, harsh chemotherapy and immunosuppression with a lower dose of radiation, immunosuppression with the drug rapamycin, and a novel monoclonal antibody. The antibody, called alemtuzumab, binds to and triggers the destruction of specific white blood cells in both recipient and donor that would otherwise cause either rejection or graft-versus-host disease.

In June 2014, Tisdale reported the results of a trial using this procedure² in sickle-cell patients between 16 and 65 years of age. Of the 30 subjects, 26 experienced successful transplants and, he says, "are effectively cured". Crucially, 15 of the patients with the highest level of engrafted donor cells were able to come off immunosuppressant drugs altogether, despite retaining a significant proportion of their original immune system, indicating that they had developed long-term immune tolerance of the graft. "This is one of our most exciting findings," Tisdale says.

DEEPENING THE DONOR POOL

Although Tisdale's work has widened the potential pool of patients by removing one of the biggest limiting factors — a patient's age — the stem-cell donations still had to come from the same small group of fully matched donors. Expanding this to half-matched donors would greatly increase the number of transplants that could be performed, as almost 90% of patients have such a donor in their immediate family.

For several years, haematologist Robert Brodsky and colleagues at Johns Hopkins University School of Medicine in Baltimore, Maryland, have been developing their own protocol that allows the use of half-matched transplants in adults, again without full destruction of the recipient's bone marrow³. Brodsky's procedure does not include antibody treatment but does use radiation and immunosuppressive drugs, including rapamycin. It also uses cyclophosphamide, a long-established chemotherapeutic agent. Brodsky explains that cyclophosphamide does not harm the engrafted stem cells, as these produce an enzyme called aldehyde dehydrogenase that



Blood from a patient with sickle-cell disease before (top) and after (bottom) a stem-cell transplant.

renders them resistant to it. But the lymphocytes that cause graft-versus-host disease have very little of the enzyme and so are sensitive to the drug.

The main problem is getting the graft to take, says Brodsky, who had an initial success rate of only around 60% because of graft rejection problems. By modifying the procedure he has increased this to nearly 75%, still lower than the 87% achieved by Tisdale using fully matched donors. But the lower success rate is outweighed by a tenfold increase in the donor pool, he says, bringing a potential sickle-cell cure to many more patients.

Brodsky reports that other teams are following his lead, and says the more gentle procedure using post-transplant cyclophosphamide is spreading across the United States and overseas. But the numbers treated still remain low — in all, maybe 40 to 50 people worldwide — and an important part of getting more patients treated is educating the wider medical community. "Even in the United States, most patients are not seen in specialist centres," Brodsky says, and non-specialist doctors tend to be unaware that transplantation is becoming an option for adults.

Mickey Koh, director of stem-cell transplantation at St George's Hospital in London, is using a modified version of Tisdale's techniques. "Our first adult patient was treated in early 2011 and is now leading a normal life with no drugs required and no evidence of graft-versus-host disease," Koh reports. If such positive results continue to accumulate, he believes that stem-cell transplantation could soon become a standard treatment option for adults.

BACK TO THE WOMB

The most dramatic extension of the age range for stem-cell transplants, however, takes it all the way back to the womb, where the affected

cells are replaced before birth. A team led by Alan Flake, director of the Children's Center for Fetal Research at the Children's Hospital of Philadelphia, Pennsylvania, has successfully treated sickle-cell disease *in utero* in mice and dogs⁴. He is now running a trial in monkeys, and is applying for regulatory approval for a small-scale clinical trial with human fetal patients within the next two years.

Flake and his colleagues collect stem cells from the mother and inject them into the bloodstream of the fetus using a fine needle guided by ultrasound scanning. No destruction of the fetal immune system is required because the fetal and maternal immune systems are naturally tolerant of one another. The most recent results in dogs achieved stable mixes of cells from mother and offspring, at levels that could make a therapeutic difference, in 40% of cases. More impressively, the trials also showed that the recipients were completely tolerant of the donor cells. If this can be replicated in humans, then even in cases when a graft does not take, a non-toxic follow-up transplant involving minimal bone-marrow damage could be given after birth, Flake says. The team hopes to increase the success rate for graft acceptance. But even if it remains at 40%, he says, "the only major risk of failure is that the child will still be born with sickle-cell disease, which would have happened anyway." His institution's discussions with the sickle-cell community have indicated that many women carrying an affected fetus would consider a fetal transplant if it were available.

These are all promising developments, but Brodsky and Koh raise issues that are often forgotten when sophisticated treatments are pioneered in developed nations. "Most sickle-cell patients are in Africa, where few countries have the resources to undertake these types of transplant," Brodsky says. He estimates that a transplant at his centre would cost around US\$300,000 — a big upfront expense, even though it may well be recouped by avoiding the costly long-term treatment that would otherwise be required. Some people from Africa have travelled abroad for transplants, but it's an option limited to a financially privileged few. Such issues serve as a reminder that although stem-cell transplantation is broadening its reach, many sickle-cell patients cannot yet benefit. Bringing the emerging treatments to every part of the world that needs them will be an economic and political challenge as much as a scientific one. ■

Andrew R. Scott is a freelance science writer based in Perth, UK.

1. Kanter, J. & Kruse-Jarres, R. *Blood Rev.* **27**, 279–287 (2013).
2. Hsieh, M. M. *et al. J. Am. Med. Assoc.* **312**, 48–56 (2014).
3. Bolaños-Meade, J. *et al. Blood* **120**, 4285–4291 (2012).
4. Vrecanek, J. D. *et al. Blood* <http://dx.doi.org/10.1182/blood-2013-11-537571> (2014).

PERSPECTIVE



Thinking beyond survival

Quality of life has become a vital consideration for children with sickle-cell disease in high-income countries, says **Michael R. DeBaun**.

Children born with sickle-cell disease in high-income countries are now expected to live to adulthood, an improvement in life expectancy that is a major medical advance. In just 19 years, from 1983 to 2002, the mortality rate for sickle-cell disease in children younger than four years old living in the United States dropped¹ by 68%. Adults with the disease can now plan to live until their fifth decade. Yet this transformation from lethal disease to chronic condition has led to new challenges. Children with sickle-cell disease now face long-term health problems, such as stroke and incurable organ damage, and their care is complicated by issues in the home that are unrelated to the condition, such as poverty and parental mental health.

Now that children are living longer, their parents' priorities have shifted from simple survival to long-term, quality-of-life concerns — especially because sickle cell is accompanied by a high risk of brain injury, something that can affect a child's performance in the classroom as well as future employment opportunities. Brain injury occurs when a child has a stroke, which may be the greatest fear held by young people with the disorder and their parents. Because of the tendency for sickled red blood cells to stick together and block blood vessels, blood clots in the brain are a very real risk for children with the condition. These can take the form of overt strokes, which cause familiar, outward symptoms such as paralysis, or of 'silent strokes', which are more specific to people with sickle cell and do not have immediate, obvious effects but can significantly affect behaviour and IQ. Before the end of the Stroke Prevention Trial in Sickle Cell Anemia (STOP)² in 1998, approximately 11% of children with sickle-cell disease in high-income countries developed a stroke. The STOP results showed that regular blood transfusions reduce the relative risk of overt stroke by up to 92%. Since this was implemented, overt stroke occurs in just 1% of these children. But preventing strokes using regular blood transfusions comes at a cost: red blood cells contain a lot of iron, and increasing a person's blood volume with transfusions leads to excessive iron stores that eventually require removal through monthly, lifelong chelation therapy.

Unfortunately, although blood-transfusion therapy has proved helpful in preventing overt strokes, there is still no known strategy for preventing silent strokes (although transfusions are effective at preventing further strokes³). Today, silent strokes occur in at least 30% of under-14s who have sickle-cell disease. Silent strokes are associated with loss of cognition, which can affect education and employment, and lead to increased risk of further overt and silent strokes.

In addition to its potential effect on the brain, sickle-cell disease can pose serious risks to other vital organs. In the kidneys, it can trigger a condition known as glomerular hyperfiltration disease, which can lead to kidney failure. In the lungs, the disease can cause a life-threatening condition referred to as acute chest syndrome, which causes difficulty in breathing, chest pain and decreased oxygen delivery. Other affected organs include the eyes, causing retinopathy, which

leads to blindness, and the heart, in which cardiovascular disease initiated by sickle-cell disease can cause premature death when compared to those without the condition.

ROUTINE VISITS

Such complications result in a lifelong schedule of annual visits to an array of specialists to ensure appropriate disease management, in addition to regular health check-ups. Over a year, a typical ten-year-old with sickle-cell disease who lives in a high-income country will have a visit schedule that includes: a type of ultrasound scan to their head called a transcranial Doppler to determine risk of overt stroke; three or four visits for surveillance of toxicity associated with the only approved sickle-cell medication, hydroxyurea (we recommend hydroxyurea therapy to all our patients with sickle-cell anaemia who are over five years old); annual influenza immunization delivered by the primary care provider; a dental visit as part of routine oral hygiene; an annual ophthalmology visit; and, if the child has a history of wheezing, monitoring for obstructive lung disease.

Such a demanding medical schedule is rarely met by even the most diligent, stay-at-home parent. In reality, the majority of children with sickle-cell disease in the United States come from families in which parents live hand-to-mouth, which has a considerable impact on the biology of the disease. We illustrated this in a recent study, in which we found that a silent stroke was associated with a 5-point drop in IQ, and that having a non-college educated head of household was associated with a 6-point drop⁴. And every extra \$1,000 of family income was associated with a 0.33-point increase. Any rigorous approach that evaluates the impact of silent

strokes on IQ must also evaluate non-biological factors such as the home environment.

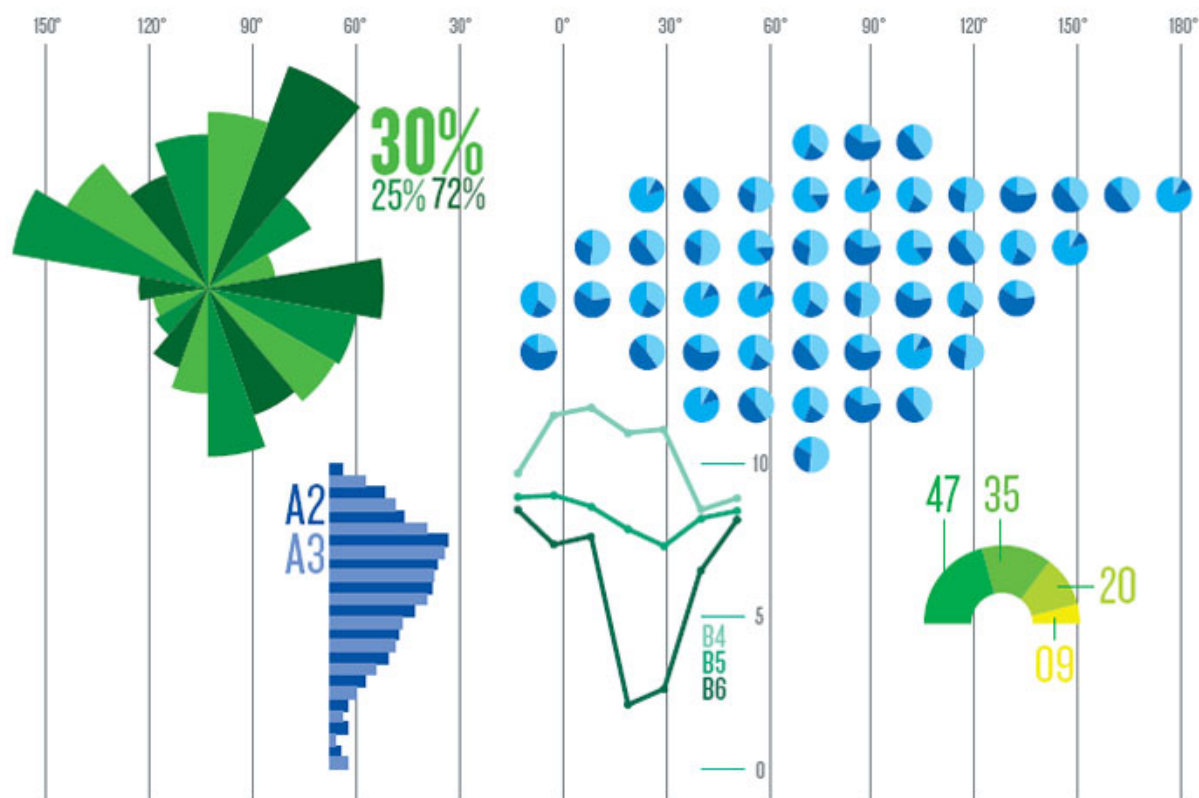
The roadmap for how to decrease childhood mortality from sickle-cell disease has been well established in high-income countries. Our current task is determining how to improve quality of life and decrease associated diseases while incorporating the context of a child's home environment, where poverty and other factors are likely to affect the disease. Both biological and non-biological risk factors must be considered collectively if we expect to not only extend the lives of children born with sickle-cell disease, but also improve their quality of life. ■

Michael R. DeBaun is director of the Vanderbilt-Meharry Center of Excellence in Sickle Cell Disease at Vanderbilt University Medical Center in Nashville, Tennessee.
e-mail: m.debaun@vanderbilt.edu

1. Yanni, E. et al. *J. Pediatr.* **54**, 541–545 (2009).
2. Adams, R. J. et al. *N. Engl. J. Med.* **339**, 5–11 (1998).
3. DeBaun, M. R. et al. *N. Engl. J. Med.* **371**, 699–710 (2014).
4. King, A. A. et al. *Am. J. Hemat.* **89**, 162–167 (2014).

nature INDEX 2014

GLOBAL



nature

INDEX 2014

GLOBAL

NATURE, VOL. 515, NO. 7526 (13 NOVEMBER 2014)

COVER ART: JASIEK KRZYSZTOFIK

This first edition of the Nature Index 2014 Global supplement provides a snapshot of results from the new Nature Index. In this supplement, we turn a spotlight on the countries and institutions around the world that contributed to some of the highest quality research over the previous calendar year.

The Nature Index provides a new way to look at the scientific literature — and to those research organizations that contribute to it. By looking at articles from only a small group of journals, most favoured by the scientific community as a place to publish their best research, we hope to provide a new level of analysis that is more targeted and hence more malleable. We want users to be able to tease out patterns of research, look at trends, analyse individual strengths, and investigate how institutions and countries collaborate. The story behind the Nature Index is outlined on page S52.

In this supplement, our analysis includes layers of information from other data sources, for example demographics, national spend on research and development, changes to science policy and funding, and even altmetrics (online and social media coverage), which help put the Nature Index data into perspective. The top level, global results can be seen on page S56.

Drilling down by region (starting with North America on page S60),

we highlight the most interesting patterns in research output in this snapshot of the Nature Index. Within each region, we try to identify national hotspots for high-quality research, based not just on output quantity but also on a range of indicators — for example, the number of researchers and the ratio of collaborators — that help put the data in context and allow a more nuanced view of these patterns.

Focusing down further, we look at some of the notable institutions in each country, and use the Nature Index data to compare and contrast between them, aiming to tease out specific institutional strengths.

Above all, our hope is that this supplement, rather than providing some authoritative analysis, will act as a conversation starter and a nucleation point for ideas for further analysis. Every reader of this supplement and user of natureindex.com will have their own specific interests and questions they want to address. We encourage use of the freely-available data to do just that. And we hope the conversation we have started here will prove useful for researchers, institutions, analysts and policy-makers alike.

Nick Campbell
Executive Editor, Nature

Michelle Grayson
Senior Editor, Nature Supplements

CONTENTS

- S52 INTRODUCING THE INDEX**
How the index criteria were established
- S56 GLOBAL OVERVIEW**
A world of achievement in graphics
- S60 NORTH AMERICA**
Strides ahead despite flatline funding
- S66 NORTH & WESTERN EUROPE**
Efficient, but struggling to collaborate
- S73 EAST & SOUTHEAST ASIA**
Budgets and quality grow within a narrow focus
- S82 CENTRAL, EAST & SOUTH EUROPE**
Physical science dominant amid shake-up
- S84 AUSTRALASIA & PACIFIC ISLANDS**
Australia does more with less
- S88 WEST ASIA**
Dynamic science in region of upheaval
- S89 CENTRAL & SOUTH ASIA**
A love affair with chemistry and physics
- S91 MIDDLE & SOUTH AMERICA**
Reaching out for international partners
- S92 AFRICA**
Heavily dependent on foreign funding
- S94 A GUIDE TO THE NATURE INDEX**
How to get the best from the index
- S98 TABLES**

EDITORIAL: Herb Brody, Michelle Grayson, Rebecca Dargie, Victoria Kitchener, Stephen Pincock, Nicola Jones, Wilson da Silva, Eric Bender, Mohammed Yahia, Subhra Priyadarshini, Linda Nordling, Hepeng Jia, Rafael Garcia, Claire Ainsworth, Barbara Casassus.
EDITORIAL SUPPORT: Nobuko Miyairi, Larissa Kogleck. **ART & DESIGN:** Wesley Fernandes, Mohamed Ashour, Alisdair Macdonald, Andrea Duffy, Chris Gilloch. **WEB & DATA:** Bob Edenbach, Olivier Lechevalier, Yuxin Wang, Naomi Nakahara, Masamichi Wada, Jyoti Miglani, Akiko Murakami, Takeshi Ouchi, Maxime Fontaine, Jennie Pao. **PRODUCTION:** Sue Gray, Karl Smart, Ian Pope, Robert Sullivan. **MARKETING:** Hannah Phipps. **SALES:** Nils Moeller, Kate Yoneyama, Kylie Ahern, Janet Cen, Yuki Fujiwara, Maria Kubalova, Krystal Trenchfield, Stella Yan. **PROJECT MANAGER:** Anastasia Panoutsou. **ART DIRECTOR:** Kelly Buckheit Krause. **PUBLISHING:** Nick Campbell, Richard Hughes, David Swinbanks.

NATURE INDEX 2014 GLOBAL

The Nature Index 2014 Global, a supplement to *Nature*, is produced by Nature Publishing Group, a division of Macmillan Publishers Ltd. This publication is based on data from the Nature Index, a website maintained by Nature Publishing Group and made freely available at natureindex.com.

Nature Editorial Offices
The Macmillan Building
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0)20 7833 4000
Fax: +44 (0)20 7843 4596/7

CUSTOMER SERVICES

To advertise with the Nature Index, please visit natureindex.com/support
feedback@nature.com
Copyright © 2014 Nature Publishing Group.
All rights reserved.

Introducing the index

The Nature Index allows us to track contributions by countries and by research institutions — academic, government and commercial — to selective scientific journals, independently chosen by active scientists. Analysis of this database provides insight into global hotspots for high-quality research.

The aim of the Nature Index is to provide an indicator of patterns of high-quality research output across the globe. At its core are 68 journals, independently chosen by researchers as being where they would want to publish their most significant research. We identify author affiliations on each paper, as well as tease out the relationships between research organizations, to allow us to track scientific output for institutions and countries. Snapshot data from the Nature Index are openly available under a Creative Commons licence, so that users can analyse scientific research outputs themselves.

The group of journals at the heart of the Nature Index is chosen under the following fundamental principles:

- The journals included are selected by a panel of active scientists, independently of the Nature Publishing Group.
- The choices reflect researchers' perception of the journals' content, rather than measures such as impact factor.

We believe that, at the time of selection, the list amounts to a reasonably consensual, upper echelon of journals in the natural sciences. It includes some of the most highly selective journals within the main disciplines of the natural sciences as well as highly selective multidisciplinary journals.

The list of 68 journals used in this initial version of the Nature Index was compiled in 2011. The journals included, and their number, will be reviewed before the next edition.

JOURNAL SELECTION

We gave prime responsibility for the selection of journals to two panels of scientists, one drawn from the physical sciences, the other from the life sciences — each headed by a chairperson. Preliminary suggestions for the chairs of the panels were made by the editorial staff of *Nature* who are involved in the peer review and selection of submitted research papers. The Editor-in-Chief of *Nature* signed off on the choice of chairs:

Chair of Physical Sciences Panel: John Morton, then at the University of Oxford, now at the London Centre for Nanotechnology and Department of Electronic and Electrical Engineering, University College London

Chair of Life Sciences Panel: Yin-Biao Sun, Randall Division of Cell and Molecular Biophysics, Kings College London

At the chairs' request, an initial list of panel members was proposed by the editorial staff of *Nature* journals. The criteria for panel members were: they be established and fully active in research (therefore more likely to be mid-career than late-career); they should be drawn from the main disciplines of natural science; they should represent all active science regions worldwide; and there should be a gender balance.

**“WE HOPE THE INDEX
WILL FIND A NICHE
AMONG THE TOOLS TO
TRACK AND QUANTIFY
RESEARCH.”**

The chairs signed off on the ultimate choice of panel members. They include 68 scientists in all, not counting the chairs.

We asked each panellist to list the journals in which they would most like to publish their best work, to a maximum of 10. They were asked to list these journals in order of preference.

To aggregate these responses, each panellist's first journal was awarded ten points, the second journal nine points, and so forth. We recorded both the total number of points that a journal accumulated and the number of panellists who voted for a journal. The chairs used these scores to analyse the popularity of each journal identified.

CONFIRMATORY SURVEY

We felt it important to obtain a broader degree of input and validation, so we conducted a large-scale survey of researchers.

We emailed 100,000 scientists in the life, physical and medical sciences with an online questionnaire. We targeted a broad geographical mix of scientists across Europe, North America, Asia and the rest of the world, receiving more than 2,800 responses from across the major disciplines of the natural sciences.

Only scientists who indicated that they have published in the past two years were included in the survey results, to ensure we polled active scientists.

There was a high degree of convergence between the panel and survey outputs for the most popular journals, with more divergence further down the list. But this process was not a number-crunching exercise. The purpose was to assist our panellists and especially the chairs in producing a final list of journals, taking all qualitative judgements and quantitative inputs into account. The final selection was entirely the responsibility of the panel chairs.

FINAL ADJUSTMENTS

The final step in this exercise was to compare our selected journals against the total output of research papers. Our aim was that the ratio of disciplines within the Nature Index should be roughly in line with annual scientific output, with no single discipline contributing to the Nature Index to an inequitable degree. If there are any gross imbalances, these should be acknowledged so that users can take them into account when assessing the patterns of high-quality research output.

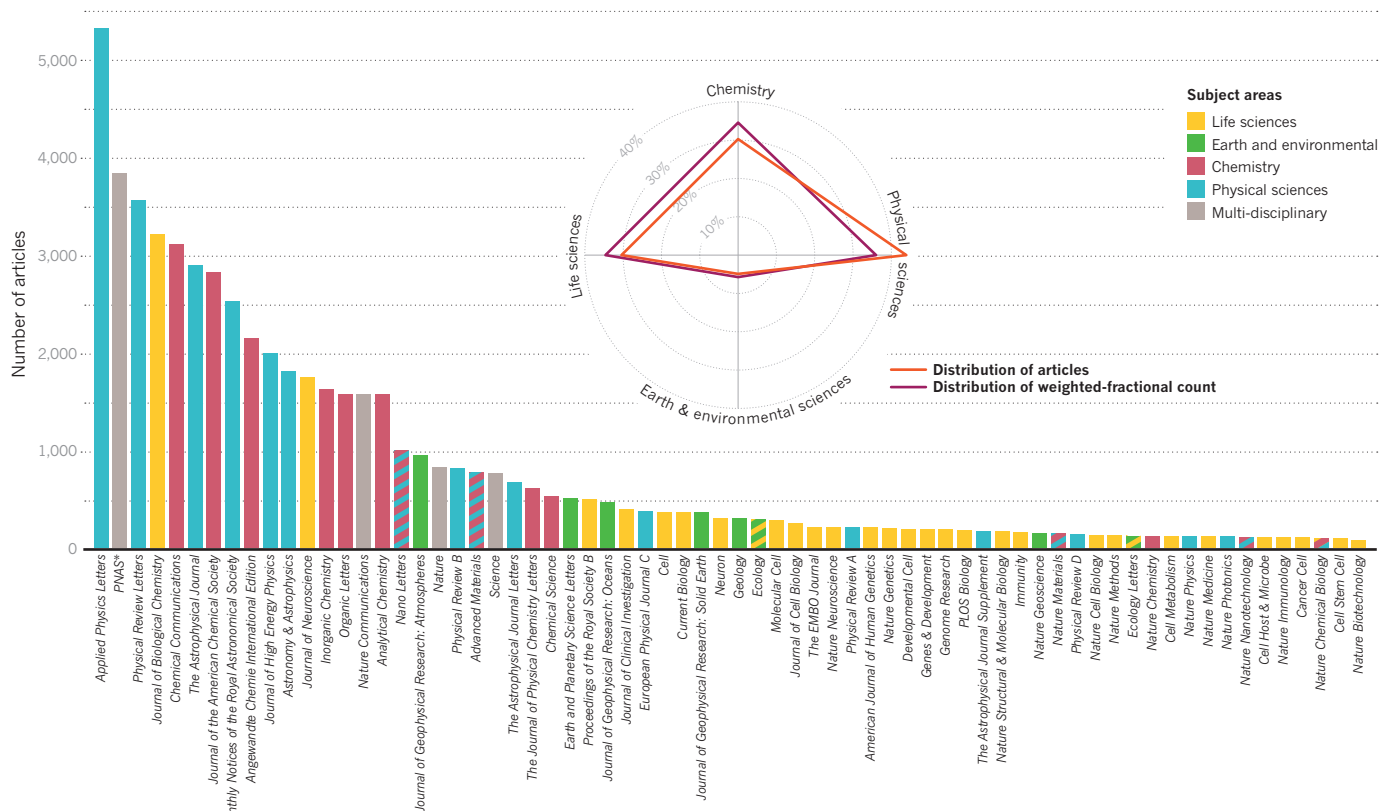
For this reason we provide three measures within the index: the raw article count (AC); the fractional count (FC), which apportions article count for each contributor; and a weighted-fractional count (WFC), which applies a weighting to correct for the one gross imbalance we identified in terms of discipline representation in the Nature Index (for more about how these are calculated, see 'A guide to the Nature Index', page S94).

That one striking imbalance was in the field of astronomy and astrophysics, where our selected journals represent about 50% of all papers published in international journals in this discipline.

This proportion was approximately five-times the equivalent figures for other fields. Therefore, although the data for astronomy and astrophysics are compiled in exactly the same way as for all other disciplines, in the WFC articles from these journals are assigned one-fifth the weight of other articles. Or to put it another way: the fractional count from those journals is multiplied by 0.2 to derive the WFC.

JOURNAL, ARTICLE AND SUBJECT DISTRIBUTION

There are 68 journals in the Nature Index. The number of articles published in 2013 and included from each journal is shown below, alongside how these map to the four main subject categories. Note: for *Physical Review A*, *Physical Review B* and *Physical Review D* only research papers in the Rapid Communications section and/or those selected to be 'Featured in Physics' (the website that spotlights the research of broad interest from the American Physical Society journal collection) were included in the Nature Index.



*Proceedings of the National Academy of Sciences of the United States of America.

This is intended to provide due representation to all fields covered by the Nature Index in any multi-disciplinary analysis of institutional outputs.

In the freely available Nature Index dataset, we leave it to the user to decide how best to use the three measures.

Clearly different measures and different subsets of the index will be more appropriate depending on the particular interests of the user. However, for the purposes of analysing and assessing global patterns within this supplement, we have, in general, focused on the WFC and the AC because they provide interesting and complementary information.

We recognize that the weighting we have applied as a broad-brush correction for the over-representation of astronomy articles is only one of many that might be applied.

More intricate and complicated weightings aimed at normalising the data for factors such as funding levels, numbers of researchers and so on, are just some of the measures that might be taken into account in assessing the patterns within the Nature Index. We encourage users to suggest and apply their own ideas for weighting the data.

The approach applied to the weighted-fractional count and all decisions regarding

the selection of journals in the Nature Index were signed off by the chairs of the journal-selection panels.

SIGNIFICANCE OF THE NATURE INDEX

The journals at the heart of the Nature Index were selected for this purpose alone. They reflect the judgment of the panellists and, ultimately, the panel chairs. This exercise was not intended to provide any absolute comparison between journals within disciplines. The process is founded on a pragmatically minded aggregation of judgments, and the lower-cut-off point is entirely subjective. There is no implication that a journal lying below that threshold is in any way inferior to all of those above it. And indeed, there is every chance that these journals will be included in future iterations of the index.

We hope that the Nature Index will find a niche among the tools that research organizations use to track and quantify research outputs and to develop comparisons across peer institutions.

Snapshot data from the Nature Index are openly available under a Creative Commons licence at natureindex.com. Data are updated monthly, and cover the most recently processed 12 months. ■

ORIGINS OF THE NATURE INDEX

In 2005, NPG began compiling the Nature Publishing Index (NPI). This counted the contribution of research institutions exclusively to *Nature* and all the *Nature* research journals. The most recent supplement, specifically for the Asia-Pacific region, was published in March 2013 and can be found at nature.asia/publishing-index-asia-pacific. As there is no need for a journal selection process, the NPI involved no external advisors or selection panels.

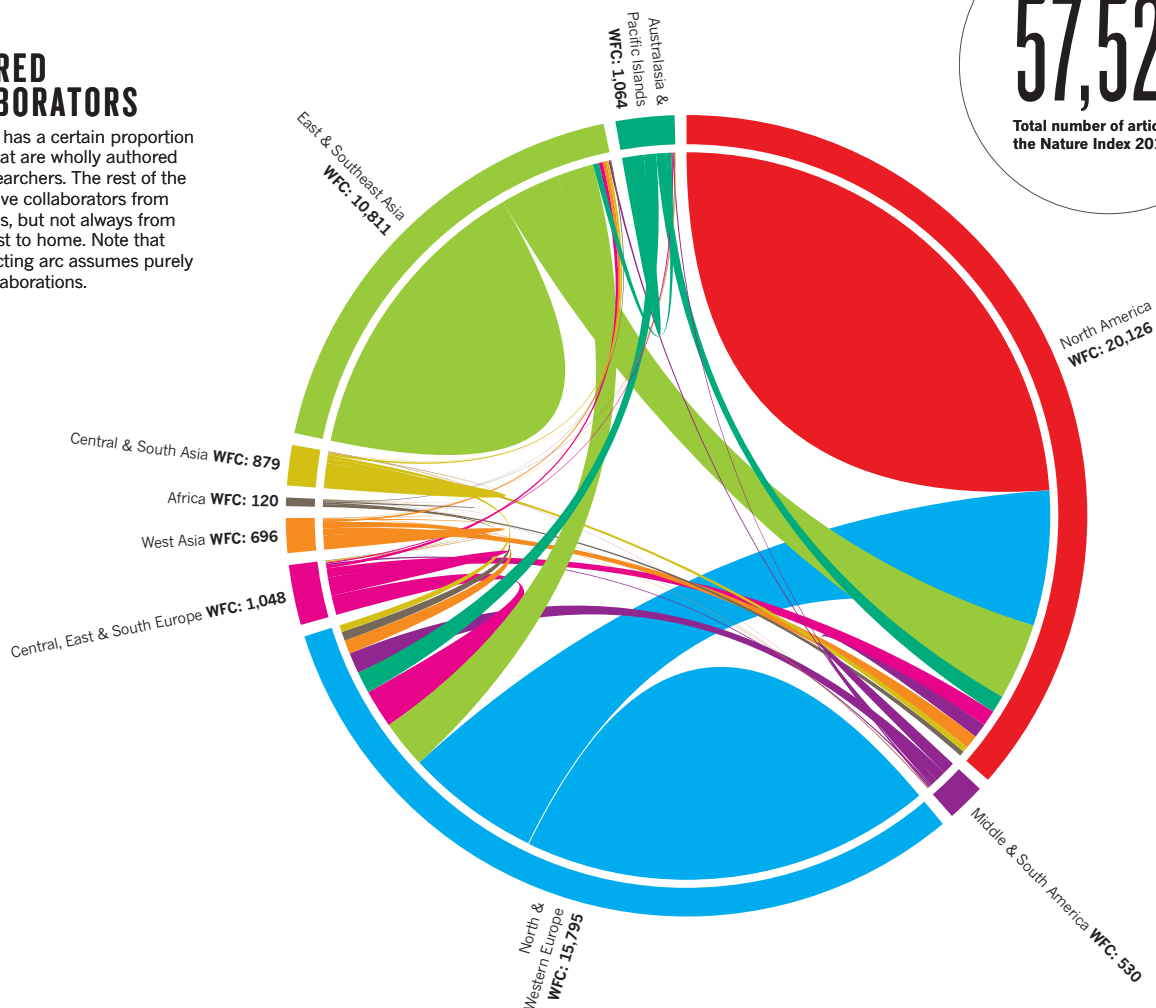
The NPI showed that this type of analysis was feasible — and moreover would be useful. The Nature Index is intended to be a more inclusive database of contributions by research institutions to the scientific literature, ranging well beyond the *Nature* journals — and not necessarily including them all (for example, the current Nature Index does not include *Nature Climate Change*, which at the time of the panel's selection exercise was too new to have established a reputation). Nevertheless, the Nature Index builds on the same principles as the NPI, albeit on a much bigger scale.

GLOBAL OVERVIEW

High-level results from the Nature Index show three strong regions, but also reveal a more nuanced picture.

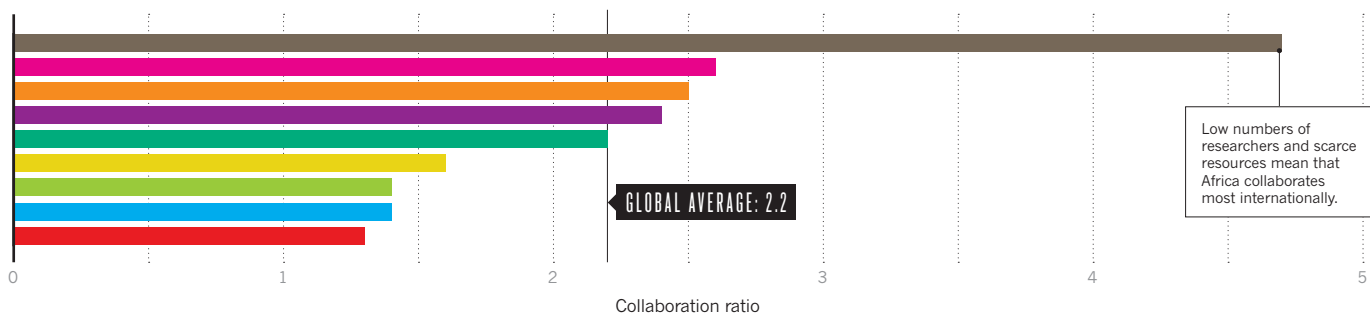
FAVOURED COLLABORATORS

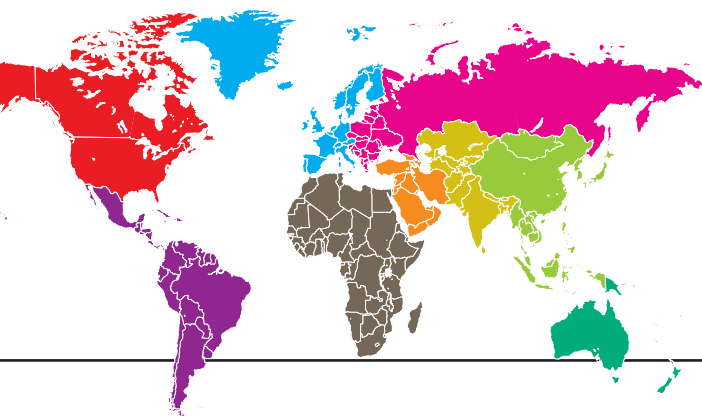
Each region has a certain proportion of papers that are wholly authored by local researchers. The rest of the papers involve collaborators from other regions, but not always from those nearest to home. Note that each connecting arc assumes purely bilateral collaborations.



FREQUENCY OF COLLABORATION

The ratio of article count (AC) to fractional count (FC) gives a measure of the region's propensity to collaborate. A higher number means more collaborators from outside the region per paper.





Regions

- North America
- Middle & South America
- North & Western Europe
- Central, East & South Europe
- West Asia
- Africa
- Central & South Asia
- East & Southeast Asia
- Australasia & Pacific Islands

Acronyms

AC: article count
(number of papers)

FC: fractional count
(apportions article count for each contributor)

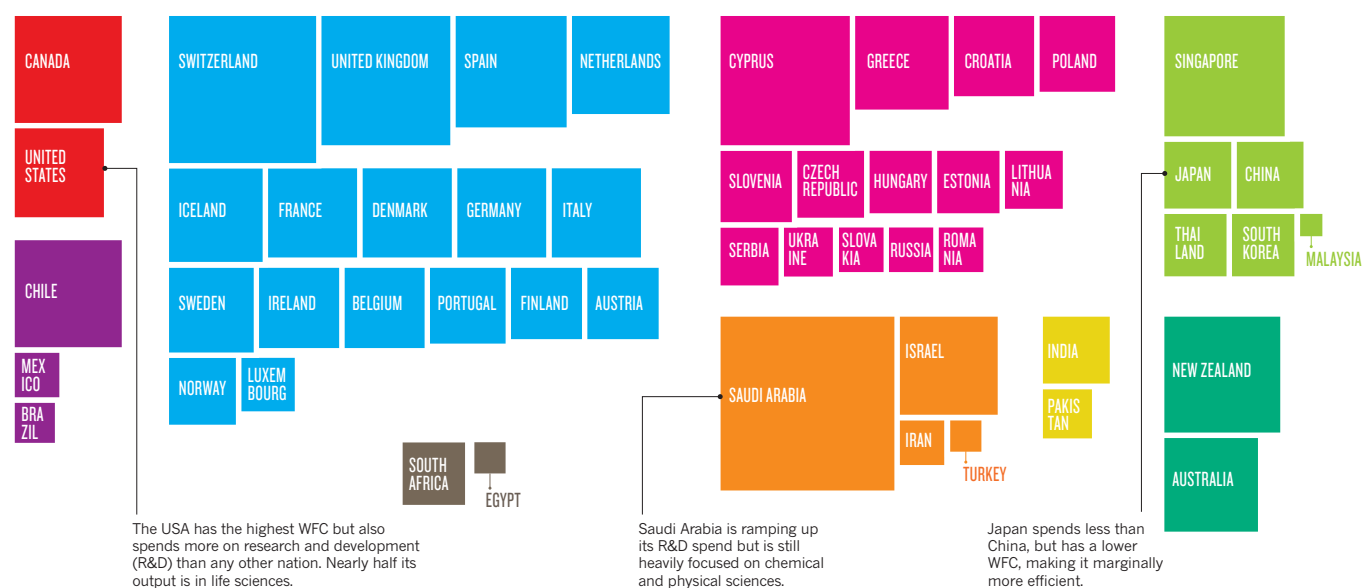
WFC: weighted fractional count
(applies a weighting to FC to correct imbalance in number of astrophysics papers)

FINANCIAL EFFICIENCY

Dividing a country's weighted fractional count (WFC) by its gross domestic expenditure on research and development (GERD, per US\$100,000, by purchasing power parity) gives a measure of its financial efficiency. So the larger the square, the higher the Nature Index output (as measured by WFC) per dollar invested.

3.42

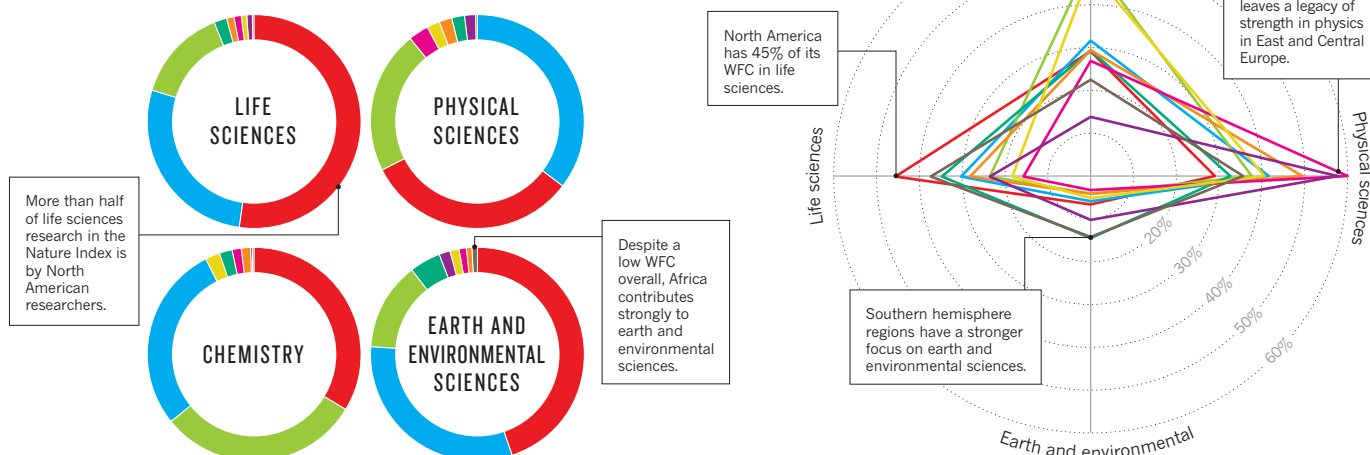
Global efficiency average



1. Only countries with a WFC>10 and GERD data from 2008 or later are included. Source: United Nations Educational, Scientific and Cultural Organization (UNESCO).

SUBJECT STRENGTHS

On a regional basis, North America has the highest WFC in three of the four subject areas. However, each region has a unique subject footprint.



North America

For many decades, North America has led the world with the breadth and depth of its science, and its continued dominance is apparent in the Nature Index results.

ARTICLE COUNT (AC): **29,325**
FRACTIONAL COUNT (FC): **22,276**
WEIGHTED FRACTIONAL COUNT (WFC): **20,126**

North America is strides ahead of all other regions, including its nearest rival North and Western Europe, in producing high-quality science. Boasting a WFC of more than 20,000 in the Nature Index, its two constituent countries, Canada and the United States, have long enjoyed strong support for natural science, mostly from their research universities. But both are adapting to significant shifts in public funding in recent years. The United States, comfortably the global leader in the index, is weathering funding cuts that have curtailed many research agencies. Canada, which comes in seventh globally and has notable strengths in life sciences and earth and environmental sciences, is attempting to tie its research more tightly to commercial innovation.

The United States dwarfs its northern neighbour on nearly every measure, but the index reveals more nuanced information about how it uses its resources. For example, according to UNESCO, there are 1.25 million researchers employed in the United States, which is 3,979 per million citizens; Canada has only 157,000 researchers, but this translates to 4,563 per million people — a higher density. The Nature Index, however, shows that the United States is better able to leverage its researchers, producing a WFC of 14.9 per thousand researchers compared to Canada's 9.4 (see 'Researcher efficiency').

Researcher efficiency may be a factor in the relative lack of collaboration with countries outside the region, a metric that is lower than the global average across all subjects. The United States is relatively self-sufficient, particularly when it comes to papers in either *Nature* or *Science* (see 'Collaboration rate').

“CHINA, OF COURSE, IS NOT THE ONLY NATION RAMPING UP R&D WHILE WE REST ON OUR LAURELS.”

The region shows an above-average contribution to the life sciences, which accounts for nearly half of its output (see 'Research strengths'). For the United States, that focus has historically been encouraged by funding from the government's medical research agency, the National Institutes of Health (NIH) — whose budget, however, is shrinking — in real terms, it is now four-fifths of its value a decade ago (National Institutes of Health). Its Canadian counterpart, the Canadian Institutes of Health Research, also has seen its purchasing power wane in recent years.

UNITED STATES: SEEKING STABLE GROWTH

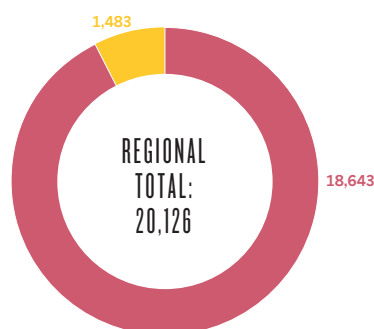
The United States' WFC in the Nature Index is 18,643, more than triple that achieved by second-place and ascendant challenger China. But maintaining this supremacy might prove difficult: 2013 budgets for many US research agencies were flat. The prospect of prolonged federal funding constraints gravely worries many in the research community. "China, of course, is not the only nation ramping up R&D while we rest on our laurels, seemingly attached to the groundless belief that the US is so ahead of other nations that we can operate on cruise control," observes Mary Woolley, president of advocacy group Research!America in Alexandria, Virginia.

The past decade has seen a drop in overall US research and development (R&D) funding of about 12% in real terms, say Matt Hourihan, director of the R&D Budget and Policy Program at the American Association for the Advancement of Science. Most of that budget decline hit defence-related R&D. "I think we will be treading water for at least a few years to come," he adds. The Obama administration has proposed near-level funding of US\$64.7 billion for basic and applied research in fiscal year 2015, and of US\$68.0 billion for developmental research.

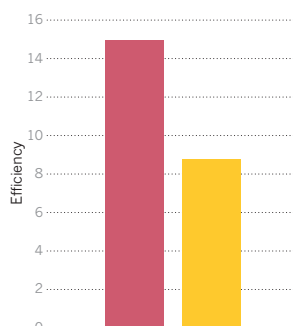
Life sciences are particularly strong in the United States. In 2013, the Nobel Prizes in

NORTH AMERICA ANALYSIS

Countries' weighted fractional count (WFC)
The United States dominates the region and leads the world.

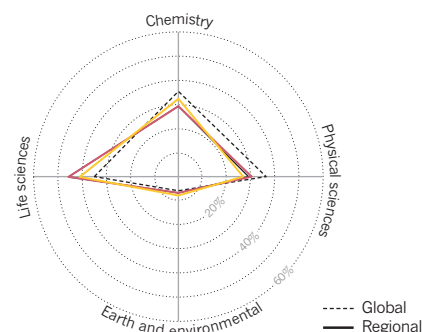


Researcher efficiency
WFC per 1,000 researchers¹.



Research strengths

Both countries are above the global average for life sciences research².



1. Source: UNESCO. 2. Subjects overlap, so the total for each country can be >100%.

Chemistry and in Physiology or Medicine both went to scientists based either wholly or partly in the United States. In the Nature Index, the country is also most dominant in the life sciences, where it accounts for just under half of the global WFC (See 'Life science share').

However, 2013 was a troubled year for US research, with federal sequester cuts in March, triggered by the failure of Congress to otherwise lower the budget deficit. These cuts hit science spending, for example slicing off about 5% of the NIH budget. Related turmoil effectively shut down most of the federal government in October, halting research at many labs. The NIH was forced to send 12,000 scientists home for "16 very, very long days," says Lawrence Tabak, NIH principal deputy director. Uncertainty about a national commitment to science "casts a pall on young people who are considering a career in biomedical research", he adds.

Overall spending has since crept up to around pre-sequester levels, but budget uncertainties continue. Another round of sequester cuts is scheduled for fiscal year 2016 unless federal deficit-reduction targets are met, says Hourihan.

CANADA: GETTING DOWN TO BUSINESS

Canada's WFC of 1,483 belies its relatively small population of 35 million (just 2 million more than Morocco), and the country "continues to punch above its weight in global science," according to Paul Dufour, an independent science and technology policy consultant based in Gatineau, Quebec. Its strength is "largely a function of the enormous expenditures since the mid-1990s in higher-education research," he says. Annual spending on natural science and engineering research by academic institutions has more than tripled in real terms, from Can\$3 billion in 1996. (However, despite the long-term upwards trend, overall federal spending on science and technology from government, industry

NATIONAL INSTITUTES OF HEALTH

Realigning for reductions



The National Cancer Institute is the NIH's biggest institute

The US National Institutes of Health (NIH) is the country's largest civilian research agency, with an annual budget of around US\$30 billion. About 10% of the budget goes on intramural research, mostly at its headquarters in Bethesda, Maryland.

When *Science* declared cancer immunotherapy as its Breakthrough of the Year in 2013, one of the two types of immunotherapy honoured was adoptive T-cell therapy, a field in which NIH researcher Steven Rosenberg played a pioneering role. That was just one mark of excellence for the NIH's intramural programme – the tenth largest contributor to high-quality scientific output in the Nature Index by WFC, and second to Harvard when just the life sciences are considered. However, much

of the NIH's work covers clinical trials, and clinical journals are not yet included in the Nature Index.

Overall, NIH published 35 papers in *Nature* and 25 in *Science* in 2013, together accounting for just over 7% of its output, with a combined WFC of 28. This places the NIH fifth in the index, just under Germany's Max Planck Society, which had about €1.53 billion (US\$1.97 billion) in public funding that year.

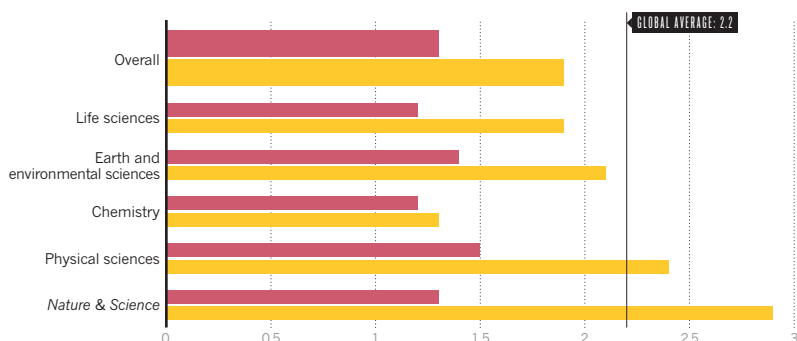
"We're seeing tremendous basic science opportunities," says Lawrence Tabak, NIH principal deputy director. However, Tabak points out that NIH overall has lost about 20% of its purchasing power in the past decade. Tabak echoes the sentiments of other observers of US science: "We need to get on a more stable trajectory."

NATIONAL CANCER INSTITUTE/SCIENCE PHOTO LIBRARY

United States Canada

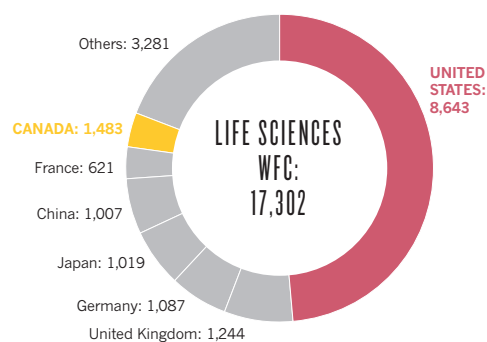
Collaboration rate

Dividing AC by FC gives a proxy for average number of collaborators per paper from outside the country.



Life science share

Nearly half of all life sciences research in the Nature Index was conducted by US scientists.





Life sciences are Harvard's strength

IT IS IN THE NATURE
AND SCIENCE COUNT
THAT HARVARD
REALLY SHOWS ITS
STRENGTH.

and academia has been heading south for several years.)

Canada spends about 1.9% of its gross domestic product (GDP) on R&D, a similar ratio to that of China and considerably less than that of the United States at 2.8%, according to estimates by Battelle Memorial Institute. R&D spending by industry and by national defence agencies, however, is much lower in Canada than in the other two countries.

Recently there has been a strong federal push to tie research more closely to business. One prominent example, Dufour notes, has been a retooling of the premiere lab, the National Research Council, to focus on business-led research. Also, "new funding going to universities has interesting clauses trying to target work closely to businesses," Dufour says.

A number of federal initiatives have strengthened research efforts. Since 1997, the Canada Foundation for Innovation programme and its partners have poured more than Can\$12 billion (about US\$11 billion) into new buildings, facilities and other research infrastructure. Beginning in 2000, the Canada Research Chairs programme has created about 2,000 research professorships, with an annual budget of about \$265 million.

Key areas of research such as quantum computing and neuroscience are supported by the related programme Canada Excellence Research Chairs, which allows Canadian universities to compete for leading international researchers. "Budgets and funding are limited, but the return on investment is worth it if Canada makes strategic investments in areas of global impact," says Feridun Hamdullahpur, president and vice-chancellor of the University of Waterloo and chair of the U15 Group of Canadian Research Universities.

The Canada Excellence Research Chairs programme seeks to maintain the country's long-standing success in attracting researchers from abroad. According to a 2014 report from The Council of Canadian Academies,

51% of individuals holding science, technology, engineering, and mathematics degrees in Canada are immigrants. (That's about twice the percentage of foreign-born college-educated scientists and engineers working in the United States, according to National Science Foundation estimates.)

THREE NORTH AMERICAN LEADERS

Three institutions exemplify the strengths of major North American academic organizations: the two US institutions that place highest in the Nature Index, Harvard University and Stanford University, and the top Canadian institution, the University of Toronto. The index reveals that these three have quite different research profiles.

Harvard has about 2,100 faculty members and its sponsored research funding totalled US\$821 million in 2013. Stanford employs roughly the same number of faculty, with a sponsored research budget of \$1.35 billion (including \$452 million for the SLAC National Accelerator Laboratory, which Stanford operates on behalf of the Department of Energy).

The University of Toronto, with a faculty of about 12,500, is the largest research organization in Canada. The university and its partner hospitals received sponsored research funding of about Can\$1.1 billion (US\$1.0 billion) for 2013. As a public institution, it has about 67,000 undergraduate students, an order of magnitude more than Harvard or Stanford.

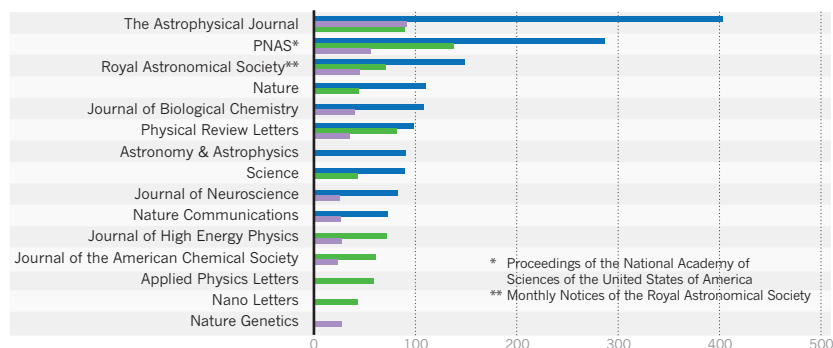
Posting a WFC of 852, Harvard comes in second among global research institutions, behind the gigantic Chinese Academy of Sciences with a WFC of 1,209. Stanford comes fifth with 553, and Toronto is 28th overall with 242.

Both Harvard and Toronto publish most in *The Astrophysical Journal*, with 403 papers from researchers at Harvard (representing 14% of all 2013 papers in this journal) and 91 from Toronto. There were also 89 papers from Stanford, although this Californian institution published most frequently in the

■ Harvard ■ Stanford ■ Toronto

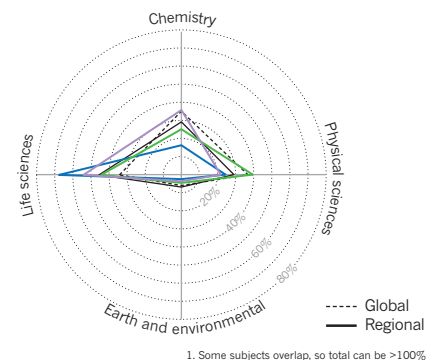
Top ten journals

Astrophysics is strong for all universities. After that, Harvard's publications are mixed; Stanford shows preference for the physical science journals, whereas Toronto favours life sciences.



Institutional subject spread

More than two-thirds of Harvard's output is in the life sciences¹.



interdisciplinary journal *Proceedings of the National Academy of Sciences* (PNAS) (see 'Top ten journals'). Nevertheless, the institution with the strongest slant towards physics overall is Stanford, with 39% of its output in this field. (see 'Institutional subject spread')

Stanford's five top journals in the Nature Index are all in the physical sciences. The university hosts the SLAC Lab, which on its own achieved a WFC of 56. Among SLAC's accomplishments was a *Nature* paper that made a significant step toward creating an electron accelerator on a chip. However, it is the Massachusetts Institute of Technology (MIT) that published more physical sciences papers than any other North American institution, with a WFC of 228 – that's more than Stanford's 215 and Harvard's 195 (see physical science table, page S107).

"CANADA CONTINUES TO PUNCH ABOVE ITS WEIGHT IN GLOBAL SCIENCE."

Earth and environmental sciences makes up only a small proportion of the total Nature Index papers, and all three institutions publish fewer than the global average in this field. The top North American institutions are two government agencies: NOAA and NASA (page S105). Stanford is fifteenth globally, but is in the process of expanding its Earth Sciences department — and in 2013 two of its faculty were given prestigious MacArthur Fellowship ("genius") awards.

One of Stanford's high-profile papers in this field was published in *Science*, and found that current climate change is happening an order of magnitude faster than at any other time in the past 65 million years.

All three institutions published most of their papers in the life sciences, particularly

Harvard where this subject accounted for more than two-thirds of its output. Indeed, Harvard is the leading institution in the Nature Index for life sciences (see life sciences table, page S104); this is the only subject where the Chinese Academy of Sciences is not top.

This achievement partly reflects the sheer size of Harvard Medical School, which has more than 10,000 academic appointments in affiliated teaching hospitals alone, compared to around 700 for Toronto and 600 for Stanford. Harvard's interdisciplinary groups, such as the Harvard Stem Cell Initiative and Wyss Institute for Biologically Inspired Engineering, also did well in *Nature* and *Science*.

It is in this count where Harvard really shows its strength. In 2013 Harvard contributed to 199 papers in total in *Nature* and *Science*, (see 'Nature and Science output') accounting for 9% of the total Nature Index articles it contributed to, making it by far the global leader by this metric. In fact there are three US institutions in the top three *Nature* and *Science* list by WFC: Harvard is followed by MIT and then Stanford (see Nature and Science table, page S108). The University of Toronto published 34 papers in *Science* or *Nature*, representing 6% of its output in the index. All three North American universities are comfortably above the world average of just over 3%. (See 'State analysis'.)

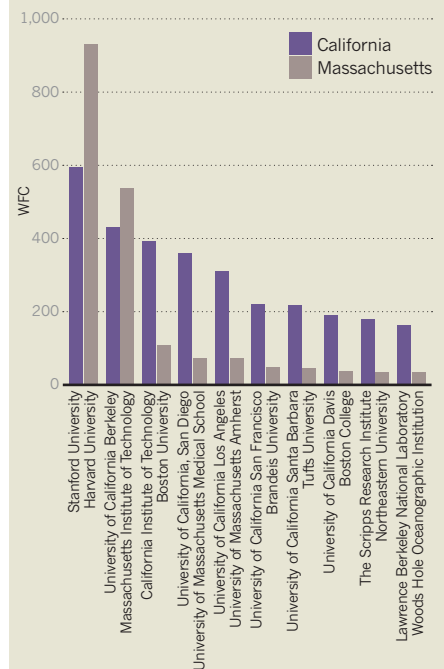
In terms of online attention (everything from Twitter to news articles) for scholarly papers, Altmetrics provides some interesting data. In this respect, Harvard has a higher visibility than the other two institutes. One of its papers from *Science*, "Poverty impedes cognitive function" (see 'Harvard's online visibility'), is in the top five papers of the year according to altmetric.com (as of 22 September 2014). However, it is a Stanford *Nature* paper, "Structural and molecular interrogation of intact biological systems", about a way to make biological tissue transparent, that gained the highest score for papers from a single institution. ■

STATE ANALYSIS

Massachusetts v. California

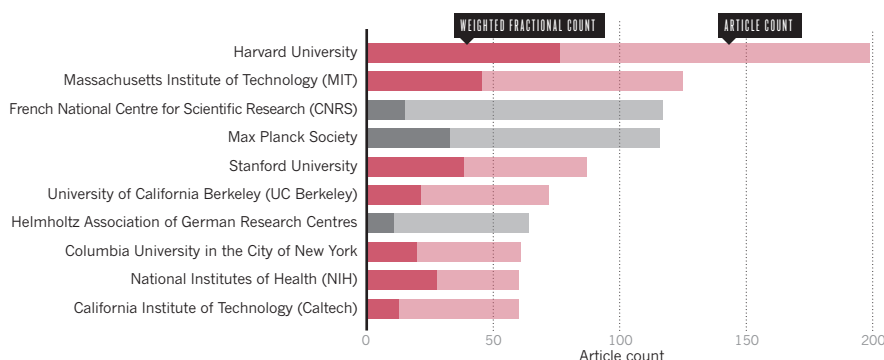
While Harvard might be dominant on an individual institution basis, it is Stanford's home state of California that shows the most scientific muscle. In the Nature Index there are 159 research institutions in California that contribute to Nature Index papers (including 12 separate campuses of the University of California) compared to only 64 in Massachusetts. After Harvard and MIT there is a big drop-off for the northeastern state, whereas California's top institutions are still delivering strong WFC scores down to the twentieth institution and beyond.

Harvard boosts its home state of Massachusetts, but California has a deeper research base.



Top ten for Nature and Science

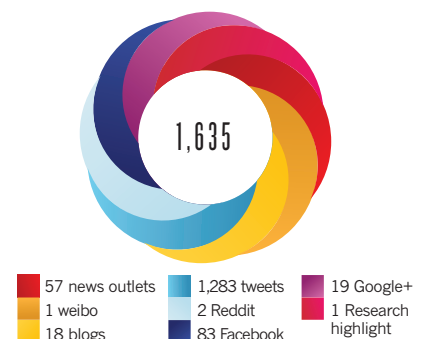
Harvard leads when just article count in these two highly selective journals are considered.



1. DOI: Science 341/6149/976. Data taken from altmetrics.com on 22 Sept 2014.

Harvard's online visibility

Harvard's most-shared paper by altmetric score is also one of the top-scoring papers¹ in all of *Science*.



North & Western Europe

A mature scientific system helps this region to produce a strong scientific output, and support for science is high – even during an economic slump. But a heavy administrative and bureaucratic burden makes collaborations difficult.

ARTICLE COUNT: **25,355**

FRACTIONAL COUNT (FC): **18,458**

WEIGHTED FRACTIONAL COUNT (WFC): **15,795**

North and western Europe is second to North America overall in the index, and is ahead in terms of the number of physical sciences papers.

Germany leads the region by article count (AC) and weighted fractional count (WFC) – a measure that adjusts AC to reflect the relative contribution of each country or institution – and is also the country that spends the most of its gross domestic product (GDP) on research and development (R&D). Over the last decade, this percentage has increased from 2.5% in 2005, to just over 2.9% in 2012. Germany's Chancellor Angela Merkel, herself a physicist, came to power in 2005 and is widely credited with pushing these priorities. During this time, the government's spending on science has risen by 60% to €14.4 billion (US\$18.5 billion), including annual increases for the large German science organizations such as the Max Planck Society (MPS).

Germany's spend as a fraction of GDP is above the average of 1.97% for the 28 European Union (EU) countries, including 2.26% for France and 1.72% for the United Kingdom, according to the Organisation for Economic Cooperation and Development (OECD). "Most EU countries have increased their R&D to GDP spending in the last 10 years, although at a slower rate since the 2008–09 economic crisis," says Dominique Guellec, head of country

studies and outlook at the OECD's Science, Technology and Innovation directorate. "The one exception is the United Kingdom, where the ratio has stagnated over the past decade."

Lidia Borrell-Damián, Director for Research and Innovation at the European University Association (EUA) in Brussels, says that Europe's R&D budgets at a regional level may be on a par with North America and East Asia, but the disparity of investment between European countries creates a problem. "Some are investing a lot, and others are investing almost nothing," she says.

The EU is trying to spur R&D spending with its Horizon 2020 Research and Innovation programme, which will provide nearly €80 billion (US\$100 billion) in funding between 2014 and 2020. But this works out as a modest annual investment compared with the €260 billion the 28 EU countries spent on R&D in 2012, Guellec notes.

North and Western Europe's success comes partly from its ability to leverage its researcher base. In terms of efficiency (shown here as WFC per 1,000 researchers), many countries in this region are among the strongest in the world. Switzerland tops the region by an enormous margin, and is also the highest in the Nature Index (see 'Researcher efficiency'). It regularly tops global surveys of competitiveness and innovation, and its science spend as

a proportion of GDP is above the EU average.

The top universities in Switzerland are the two parts of the Swiss Federal Institutes of Technology: the Federal Institute of Technology Zurich (ETH) and the Federal Polytechnic School of Lausanne (EPFL), which come sixth and eighth in the region respectively. Philippe Gillet, provost of EPFL, outlines factors that contribute to Switzerland's success: unlike France, for example, establishments in Switzerland have autonomy and total freedom over scientific and education policy; higher education, research and innovation are a priority for the country; and part of the polytechnics' mission is to help spur the economy.

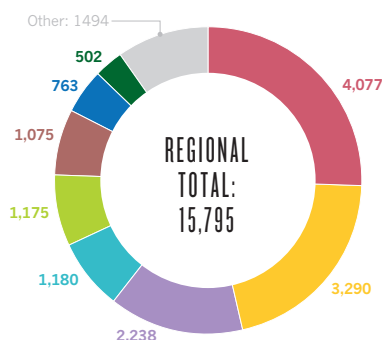
Gillet moved from his native France to Switzerland specifically to experience the system "which works". Many countries, such as France, proclaim that higher education and research are a priority, he says, but "they do not follow through with enough resources and structural reform".

Scientists in North and Western Europe are less likely to collaborate with researchers from other regions than the global average (see 'Collaboration rate'). The breadth of languages (23 official ones) in the EU is not necessarily the culprit in lack of scientific collaborations, says Bernard Meunier, vice president of the French Academy of Sciences. He says that it is very much determined by discipline. "In particle

NORTH & WESTERN EUROPE ANALYSIS

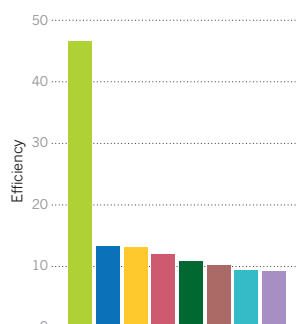
Countries' weighted fractional count (WFC)

There are 23 countries in this region, and all but 4 have a WFC of at least 1.



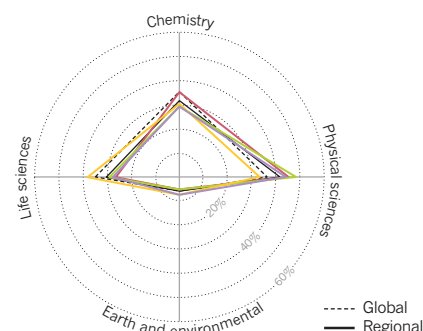
Researcher efficiency

WFC per 1,000 researchers¹.



Research strengths

The UK is strongest in the life sciences, whereas mainland Europe focuses more on physical sciences².



1. Source: UNESCO. 2. Subjects overlap, so the total for each country can be >100%.

physics, about 500 scientists collaborate in projects at the European Organization for Nuclear Research (CERN), whereas in biology, chemistry and pharmacology, research teams are in competition with each other," he explains. "Astrophysicists work with colleagues from around the world, but mathematicians lead a solitary life."

Another reason for the low level of collaboration in the EU could be the heavy administrative burden for researchers navigating different funding structures. The red tape can discourage researchers from looking abroad for collaborators, says the EU's Borrell-Damián. At a recent university conference in Italy, one rector said the research projects in progress at his university involved 150 different funding regimes, which created huge management problems. "I am sure he is not alone in this, and that large universities would have even more than 150," Borrell-Damián says. "The administrative requirements should be rationalized."

Increasing collaboration is an attractive goal, given that benefits include enriching scientists' careers, raising the profile of universities and conducting research in the best conditions possible. International cooperation means that "universities are competitive worldwide and are at the forefront of their field and even driving it," she says. That is "good for researchers, good for institutions and good for the advancement of science."

SUBJECTIVE SPREAD

Of the countries in this region, the United Kingdom is most heavily skewed towards life sciences, which comprises a third of its output in the index. On its own, the United Kingdom contributes to 7% of all the life sciences research in the index (the largest after the United States). Of the seven European institutions in the top 50 globally for life sciences, three are in the United Kingdom, led by the University of Cambridge (see life sciences table, page S104).

Duncan Maskell, Head of the School of Biological Sciences at the University of Cambridge,

says that the UK's strength in this field is the result of a supportive political and financial environment. "The government appears to understand the contribution that science – and in particular the life sciences – makes to people, and currently fosters an environment conducive to research. It's encouraging to see that it now even has a minister focused entirely on the life sciences, for example."

Germany's strength lies in the physical sciences, where it makes up 10% of the Nature Index — on par with China. Its largest research institution, the MPS, is second only to the Chinese Academy of Sciences in WFC in this field, and MPS has overall contributed to more papers in the Nature Index.

Christoph Ettl, senior scientist in the Presidential Division of the MPS, says Germany has a very strong history of physics, with Max Planck, Albert Einstein and Werner Heisenberg among others having contributed to the bedrock of quantum mechanics. Until 1976, physicists born in Germany topped the list of Nobel laureates. And the discipline has many applications beyond science, says Ettl — who is also a physicist. "Physicists have the quantitative tools for simulation, applicable to almost any field from society to the stock market."

INSTITUTIONAL COMPARISON

In most European countries, the production of papers is spread between many institutions. In the United Kingdom, universities — or tertiary educational institutions — produce the bulk of the country's research in the index; in Germany, it tends to be research-only societies and associations that dominate, including MPS, and the Helmholtz and Leibniz Associations.

The story is very different in France, where the National Centre for Scientific Research (CNRS) towers above all other institutions with a WFC of 719 — similar to Germany's MPS at 726 (see France). Then there is a notable drop-off: the next French research institution, the Atomic Energy and Alternative Energies Commission, has less than a fifth of this score, and

the highest placed university is the Pierre and Marie Curie University (UPMC) in Paris, with a WFC of 138 (see 'Institutional subject spread').

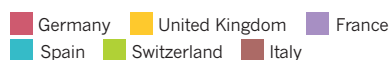
Physicist Paul Indelicato, President of Research and Innovation at UPMC, does not believe the enormity of the CNRS has a negative impact on his or other research institutions. "On the contrary, it is a strength," he says.

Comparing like-for-like single institutions across the three leading Northwest European countries pits Germany's Ludwig Maximilian University of Munich (LMU) against UPMC and the University of Cambridge.

UPMC has strong output in the physical sciences (see 'Institutional subject spread'), particularly in space sciences: more than a third of its 1,319 papers in the Nature Index are in astrophysics journals. Indeed, astronomy is a historical legacy at UPMC through its long-standing links with the Paris Observatory. UPMC founded its Paris Institute of Astrophysics in the 1930s, which a few years ago became a joint lab with the CNRS. However, because of the down-weighting of astrophysics journals in the Nature Index (see 'Introducing the index', page S52), UPMC ends up with a WFC of only 95. UPMC's namesakes — Pierre and Marie Curie — are of course famous for research in the physical sciences. "The spirit in which they conducted research still has an important influence on how we work today," says Indelicato. "We are extremely attached to Marie Curie's name and to her demand for quality research." Since the 1960s, the UPMC's mixed labs have produced three Nobel laureates in quantum physics.

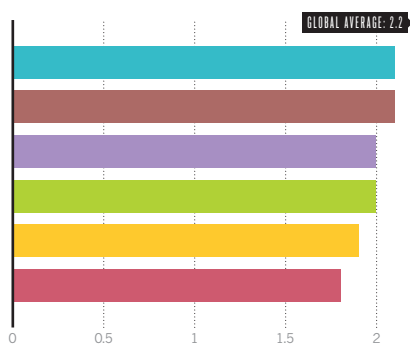
LMU, in contrast, has a more even distribution of publications across several journals, which translates into an even subject split between chemistry, physics and biology, each of which comprises roughly a third of its WFC (overlap between the subjects means that the total can be more than 100%).

Of the three, Cambridge is the strongest in the life sciences, with a WFC of 151, three-times that of LMU's. It also leans marginally more towards this field in its overall subject



Collaboration rate

Dividing AC by FC gives a proxy for average number of external collaborators per paper.

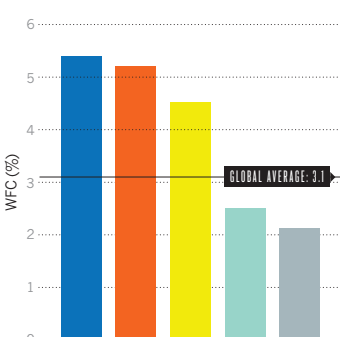


*Ludwig Maximilian University of Munich **Pierre and Marie Curie University ***French National Centre for Scientific Research. 1. Some subjects overlap, so total can be >100%.



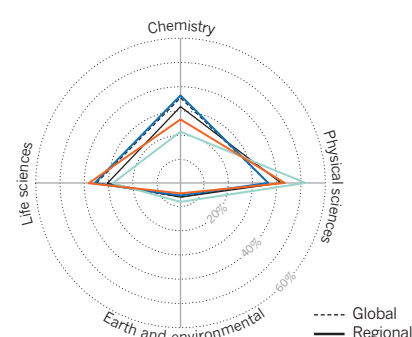
Nature and Science split

Cambridge publishes most in these two journals, but its average percentage is similar to LMU's.



Institutional subject spread

Just over half of UPMC's output is in the physical sciences, whereas the others have a more even split¹.



spread. A lot of this strength comes from neuroscience: the *Journal of Neuroscience* is one of the university's top journals by WFC. Cambridge's Maskell identifies the university's cross-disciplinary activities, including its Strategic Research Initiatives (one of which is in neuroscience), as helping to make it a leader in this field. "We have excellent opportunities to go from very basic research right through to translation, across the spectrum from areas like agricultural and conservation biology to clinical medicine." Notably, Cambridge also has the most wholly-authored papers of the three: 138 compared to 53 from LMU and none from UPMC.

When it comes to the highest impact science, both Cambridge and LMU are strong in publications in *Nature* and *Science*, well above the global average (see 'Nature and Science split').

In raw numbers, Cambridge is more successful at publishing in these two journals, but in terms of proportion of its overall output there is little to choose between the UK and German institutions.

Less-traditional measures of the impact of research, called altmetrics, track a paper's visibility in the online world including social media. All three west European institutions have contributed to papers that have scored highly, according to altmetrics.com, but it is Cambridge that appears to be the most successful. Its top paper is from 2013, involving two authors from Cambridge's Psychometrics Centre, along with a researcher from the nearby Microsoft lab. The paper seems made for social media: it describes how a variety of personal information can be gleaned by reviewing someone's Facebook 'likes' (see 'Cambridge's online visibility').

Altmetrics may be a new rating system, but Maskell sees its benefit for the university. "A high score in altmetrics reflects a high level of engagement with our research, within the sector and with the wider public. We cannot expect the public to support research if they don't know what we're doing." ■

FRANCE

Centre national de la recherche scientifique

The French National Centre for Scientific Research (CNRS) is the largest basic research agency in Europe and the largest public funding agency in France. The initial budget for 2014 stands at just under €3.3 billion (US\$4.2 billion), down from €3.4 billion in 2013. France's only multidisciplinary research agency, it employs more than 11,000 tenured researchers and 13,000 engineers and support staff. It runs 1,028 research units, more than 90% of which are joint labs with universities, specialized research organizations and industry.

The collaborative structure of the CNRS is reflected in its above-average ratio of AC to FC across all subject areas. The CNRS has more than 4,500 papers in the index, more than any other institution. But of these, publications in *Nature* and *Science* represent just 2.1% of the CNRS output, below the index average of 3.1% (see 'Nature and Science split').

The agency is often criticized for its heavyweight administration. In an interview with French economic daily *Les Echos* two years ago, Bernard Meunier, vice-president of the French Academy of Sciences, said that the number of administrative posts in the CNRS had increased nine-fold since the early 1960s, while the number of researchers had increased only four-fold. "The administration is in the process of stifling research," Meunier said. "Researchers spend their day filling out administrative documents to the detriment of their real work."

The CNRS and Germany's Max Planck Society (MPS), which have similar WFCs in the index, are both large, mainly or entirely publicly funded and offer tenured jobs. But there the similarities end, says Christoph Ettl, Senior Scientist in the Presidential Division

of the MPS. The CNRS budget is double Max Planck's, but MPS is independent at all levels, with neither the government or its top managers having a say in the work its scientists undertake or in the appointment of the society's president, he notes.

In contrast, the CNRS president is appointed by the government. The agency's scientists have freedom over their work, but are poorly paid and hold on to their jobs even if they are not productive because of civil servants' sacrosanct status in France, an economic



analyst said. "MPS keeps only the productive researchers and pays them well. But, there is neither stick nor carrot in France".

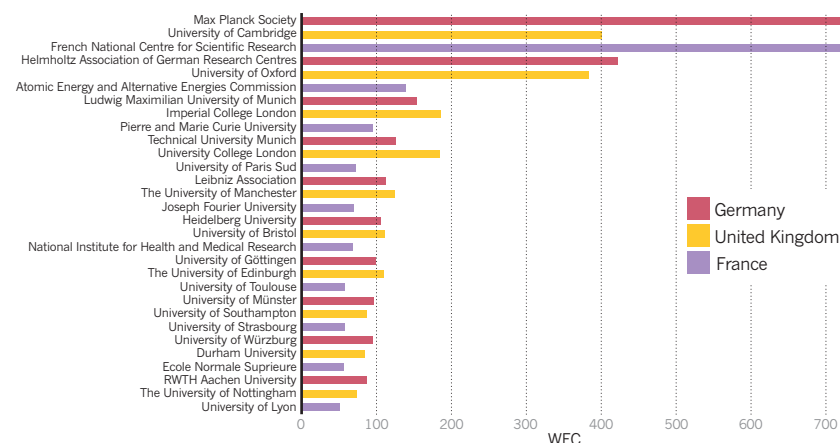
CNRS now spends about 85% of its budget on salaries, whereas MPS spends only around 40%, says Ettl and therefore has much more money available for its 80-plus research institutes. Meunier says this is a problem that is widespread in France. "Salaries should account for no more than 65% of the total," he says.

Another issue, Meunier adds, is that France concentrates too much on research structures rather than individuals. "In basic research, the government should follow the example of the European Research Council's Principal Investigator by selecting lead scientists of quality rather than focusing on institutions or labs." Many of these issues have so frustrated French researchers, that several thousand of them recently marched to Paris from many parts of the country in protest.

CNRS PHOTO THÉRIE/NICOLE TIGET

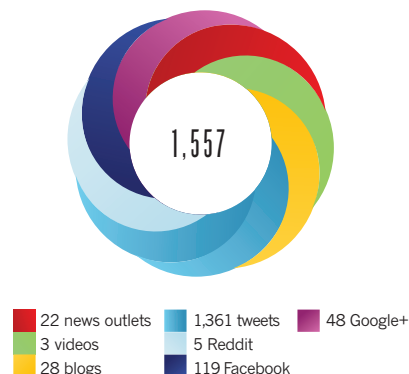
Institutional spread for top three countries

The Max Planck Society and France's CNRS dominate, but the UK has more medium-strength institutions.



Cambridge's online visibility

This paper³, analysing people's patterns of Facebook 'likes', was one of the highest scoring on altmetrics.com in 2013.



3. DOI: 10.1073/pnas.1218772110. Data taken from altmetrics.com on 22 Sept 2014.

East & Southeast Asia

China, Japan, Singapore and South Korea vividly demonstrate how significant investment in science can help to fuel national economic growth.

ARTICLE COUNT (AC): **15,638**
 FRACTIONAL COUNT (FC): **11,449**
 WEIGHTED FRACTIONAL COUNT (WFC): **10,811**

Nations across East and Southeast Asia have enjoyed success from implementing policies that increase research and development (R&D) budgets and stimulate scientific endeavours, says Tateo Arimoto, the director of Science, Technology and Innovation Policy Program at Japan's National Graduate Institute for Policy Studies. "A remarkable increase in scientific output in recent years reflects these efforts," he says. The region might be third overall by weighted fractional count (WFC), which gives a measure of the relative contribution to each paper; but three of its countries — China, Japan and South Korea — are in the top ten, led by China with a WFC of 5,206.

China and Japan, the world's second and third largest research spenders, are expected to jointly invest nearly 28% of the total US\$1.6 trillion (valued in purchasing power parity) for R&D around the world in 2014, according to US thinktank the Battelle Institute. South Korea is expected to come fifth, with a US\$63 billion research and development expenditure.

These countries are traditionally strong in physical science and chemistry — of China's total WFC of 5,206, for example, a WFC of 4,696 (or 90%) is from chemical and physical sciences. The Asian powers are spending heavily to maintain this lead, particularly in target areas such as nanoscale research. But,

Buhm Soon Park, head of the Graduate School and of science and technology policy at Korea Advanced Institute of Science & Technology (KAIST), questions whether a narrow focus is the best way forward. He says this approach will "raise the real question of how far China and Korea can go with such intensive funding".

"IT IS VERY CLEAR THAT CHINESE SCIENTISTS ARE MAKING TREMENDOUS BREAKTHROUGHS."

Soon Park's worry is shared by Fengchao Liu, a professor of science policy at Dalian University of Technology (DUT) in China, who adds that the Asian countries lag behind the United States and much of Europe in terms of generating novel research.

CHINA, THE BIG SPENDER

There's no doubt that China has the greatest growth rate, both in terms of R&D investment and scientific outputs, of the region's countries. According to UNESCO, China's research spending passed 1 trillion yuan (US\$163

billion) in 2012, or 1.98% of its GDP, surpassing the European Union's 1.96%. A decade previously, the country was investing only 1% of GDP.

From 2012 to 2013, China's WFC increased by 15% — even more impressive when compared to an increase of less than 1%, or even a decrease, for all other top-ten WFC countries.

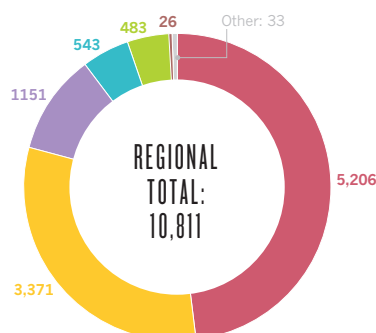
A recurring criticism is that this rise in quantity has not been matched with improved quality, but that might be changing. China also published more in *Nature* and *Science* in 2013 than in 2012, rising from a position of eight to five overall on this measure. And, says Zheng Liang of the China Institute for S&T Policy at Beijing's Tsinghua University, there's more to come. "It is very clear that Chinese scientists are making tremendous breakthroughs in areas like physical sciences," he says. "And China's capacity in technological development and engineering projects has dramatically grown."

In 2012 a Chinese paper made *Science* magazine's prestigious top-10 breakthroughs for the first time in the list's 18-year history. The study, by researchers with the Chinese Academy of Sciences (CAS) at the Daya Bay Nuclear Power Plant, measured the last unknown parameter needed to describe how neutrinos collide at near-light speed. Results are expected to help explain why the universe contains so much matter and so little antimatter.

EAST & SOUTHEAST ASIA ANALYSIS

Countries' weighted fractional count (WFC)

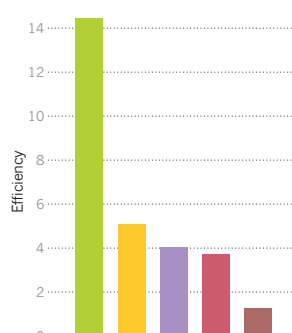
Of the 15 countries in the region, three dominate the scientific output, led by China.



1. Source: UNESCO. 2. Subjects overlap, so the total for each country can be >100%.

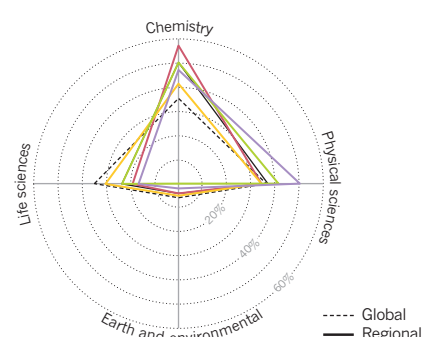
Researcher efficiency

WFC per 1,000 researchers¹.



Research strengths

Most countries focus on chemistry and physical sciences. Japan is strongest in the life sciences².



SINGAPORE

Punching above its weight



A*STAR is one of three main institutions

Although Singapore is the smallest nation in Southeast Asia, in some respects it outperforms other countries in the region and further afield. For example, the Global Competitiveness Report 2014–2015 released by the World Economic Forum in early September 2014 endowed the silver medal of competitiveness to Singapore, second only to Switzerland. This measure aims to quantify a country's drivers of productivity and prosperity.

The Nature Index adds another credential to those plaudits, showing that the nation has world-class efficiency. The rate of WFC per 1,000 researchers is 14.2 — comparable to the United States and ahead of most major European nations except Switzerland (see Researcher efficiency). According to Wong Poh-Kam, a professor of innovation policy at National University of Singapore (NUS), the city state's active international collaborations, particularly with China, contribute to its excellence.

There are three main players in Singapore: the Agency for Science, Technology and Research (A*STAR), Nanyang Technological University (NTU), and the NUS. Although chemistry is still the greatest strength for all three, there is a growing focus on government funding for life sciences.

Although China and its leading institutions are showcasing growth, its foundations could be a lot better, says Cong Cao, a leading China science policy expert at the United Kingdom's University of Nottingham, who has repeatedly urged an overhaul of China's science system. "Funding for basic research has been below 5% of total R&D spending for a long time, weakening China's innovation capacity," Cao says. Comparative figures for the United States and Japan are 19% and 12.5%, respectively. "In addition, funding is unevenly concentrated in some top academicians, resulting in reported misuses and insufficient resources for young scientists to develop."



One of Nanyang Technological University's schools

PEKING UNIVERSITY

CAS is the leading institution in the world in terms of overall publications in high-quality journals, with a WFC of 1,209 in the Nature Index and a size and stature to match: the Academy has 50,000 scientists at more than 100 research institutes. But China's leading single institution is Peking University (PKU). Like the country in general, PKU is strong in chemistry (see 'Institutional subject spread'), which chemist Song Gao, vice-president of the university, attributes partly to its strong collaborations with other institutions, such as the

CAS and the University of California.

However, when it comes to publishing in *Science* and *Nature*, PKU is on a par with CAS — but Gao feels that its scientists could do better (see 'Nature and Science split'). Gao says the university concentrates too much on sluggish research fields, and suffers from a paucity of long-term systematic studies and inadequate funding. Targeting the crossover area of chemical biology could help to boost higher-profile results, he says.

Yi Rao, former dean of the Peking University College of Life Science, agrees that PKU's life science is weak. Other universities have much bigger budgets and a longer track record in life science, he says. However, PKU has recently hired at least 30 new junior faculty members to boost research in this area. "Once our newly recruited faculty members grow up, things will change," Rao predicts.

JAPAN, PULLING OUT OF STAGNATION

Home to the top two universities in the region, the universities of Tokyo and Kyoto, Japan also enjoys a higher ratio of publications in top journals *Nature* and *Science* than other countries in the region — although at 1.9% this is still below the global average of 3.1%. Japan is also much stronger in life sciences than China and South Korea. Since 2000, 14 Japanese scientists have been awarded Nobel Prizes. The most recent Japanese-born laureates, who jointly collected the 2014 prize in physics, are Isamu Akasaki, Hiroshi Amano and Shuji Nakamura (now a US citizen) for their invention of efficient blue light-emitting diodes that enabled bright and energy-saving white light sources.

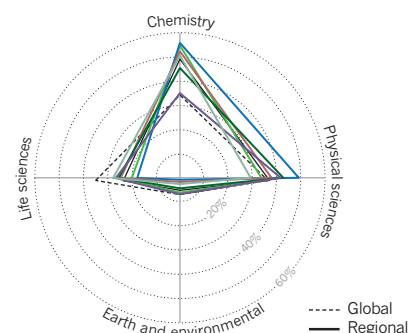
Greater funding for basic science also distinguishes Japan from China and South Korea, says KAIST's Soon Park. He says that science in these latter two countries is often perceived in "utilitarian terms" — that is, as a necessary step in the production of technology.

After years of stagnation, Japan's R&D expenditure started to grow again in 2012, rising 1.6% over the previous year to 17.4 trillion

Chinese Academy of Sciences (CAS) Korea Advanced Institute of Science and Technology (KAIST) Kyoto University
Peking University (PKU) Seoul National University (SNU) The University of Tokyo

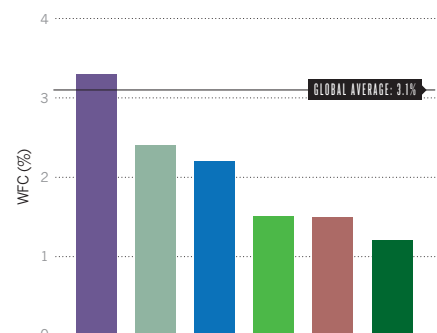
Institutional subject spread

Leading institutions in each country focus heavily on chemistry and physical sciences¹.



Nature and Science split

Within the region, only The University of Tokyo publishes above the global average.



1. Some subjects overlap, so total can be >100%.

yen (US\$165 billion). In 2013, the Council for Science and Technology Policy, chaired by Prime Minister Shinzo Abe, called for a dramatic increase in the numbers of female and



KAIST has strong international links

foreign researchers at Japanese research institutes from 3.9% to 20% by 2020. It also called for total government R&D investment to be secured at 1% of GDP.

This national ambition is expected to benefit top universities like the University of Tokyo, which is the region's biggest producer of papers in *Science* and *Nature* (making up 3.3% of its overall output in the index). One *Nature* paper, on which four out of five authors were from the University of Tokyo, considered the quantum teleportation of information or qubits, which scored well in the online and

social media-focused altmetrics system (see The University of Tokyo's online visibility).

The University of Tokyo is stronger in life science than many other institutions in the region, with 27% of its papers in this field. In the index, the university's recent highlights include a paper in *Nature* about a revolutionary method to probe molecular structure using 'crystalline sponges', and one in the *Journal of Clinical Investigation* presenting evidence that RNA splicing is involved in certain types of cancer that affect blood cell production. Such successes are proving alluring. "Promising young researchers have been moving from physics and chemistry to medical and biological sciences," says Arimoto.

"AN EXPERIMENTAL SPIRIT IS FIRMLY INGRAINED IN KAIST, SOUTH KOREA'S ELITE SCHOOL."

But, says Masatoshi Tagawa, a molecular biologist at Japan's Chiba University, the University of Tokyo, and Japan in general, risks losing its edge unless it collaborates more. In the index, Japan has one of the lowest rates of collaboration in the region (see 'Collaborative potential'). Liu of DUT agrees: "Japanese industries have rapidly expanded their collaborations with worldwide academics, but their universities are relatively conservative in both international collaborations and welcoming foreign researchers."

IMPORTANCE OF IMPACT

Although South Korea is perhaps best known for the commercial products pumped out by Samsung or Hyundai (both of which appear in the index), its basic research output is also

strong. The country has 1,953 papers in the Nature Index in 2013, with a WFC of 1,151 placing it tenth worldwide.

Across most research fields, Seoul National University (SNU) and the Korea Advanced Institute of Science and Technology (KAIST) have the country's leading WFCs, with Samsung Electronics achieving the third highest output in physical sciences. Both South Korean institutions have seen a decrease in their output of high quality science since 2012, dropping in WFC terms by 11% for SNU and 18% for KAIST.

As the country's top research university, SNU has a long and prestigious history with a strong tradition of physical sciences research, borne out by the Nature Index showing 42% of its output is in this field. Nevertheless it is growing its capacity in the life sciences, and at 26% has the highest proportional output in South Korea – despite a setback when, in 2006, stem-cell researcher Woo Suk Hwang was discredited for fabricating data on cloned human embryos.

SNU has an AC twice that of KAIST, and a WFC that is 50% higher, but KAIST's researchers have a greater percentage of papers in *Nature* and *Science*: 2.2% of all output compared to 1.2% for SNU.

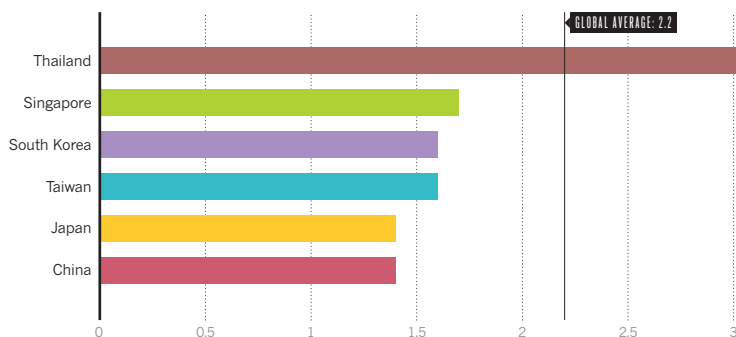
The two institutions are very different. KAIST was created in 1971 as an elite graduate school focused on research, under the Ministry of Science and Technology instead of the Ministry of Education. Its academics are better paid and have greater autonomy than at other South Korean universities, including the ability to waive tuition fees, exempt KAIST students from military service, hire new faculty members and create (or close) new academic programmes. "The experimental spirit is firmly ingrained in the school," says KAIST's Soon Park.

Sung-Mo Kang, President of KAIST, thinks international links have played a key role in its success. "Out of KAIST's total 877 faculty members recruited between 1971 and 2014, 617 professors received their doctoral degrees in the US," he says. ■

China Japan South Korea Taiwan Singapore Thailand

Collaborative potential

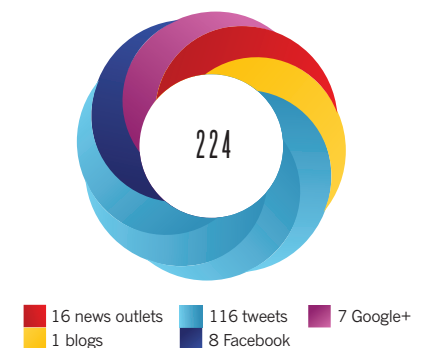
Only Thailand is above the global average for international co-authors per paper.



1. DOI: 10.1038/nature12366. Data from altmetric.com, 22 September 2014.

The University of Tokyo's online visibility

Strong coverage for its paper¹ on the quantum teleportation of information, or qubits.



Central, East & South Europe

Countries in the region are pushing to adopt a broader research base, along with updated institutional organizations. But for the most part it is the physical sciences that remain strongest.

ARTICLE COUNT: **3,351**

FRACTIONAL COUNT (FC): **1,321**

WEIGHTED FRACTIONAL COUNT (WFC): **1,048**

One of the most striking findings about this region is the subject distribution of its research footprint.

Globally, subject areas for papers in the Nature Index are fairly evenly split among the life sciences, chemistry and physical sciences, with each commanding approximately one third of the share (with earth and environment papers making up 6% of the total). In Central, East and South Europe, however, there is nearly double the output in physical sciences. To some extent, this reflects history. Science in countries in the former USSR and Eastern Bloc was affected by communist ideology: fields held to showcase the best of the Soviet system, such as space science, were favoured, while others that contradicted ideology, such as genetics, were suppressed.

As the largest country in this region, with almost 143 million people, Russia produces the most high-quality science — and, as expected, is focused mainly on the physical sciences. Its world-leading centre for nuclear physics, the Joint Institute for Nuclear Research (JINR — set up as a Soviet response to the European Organisation for Nuclear Research, also known as CERN in the 1950s) has seen its budget swell from US\$37 million in 2005 to more than \$100 million in 2010. The institute, renowned for its work on the synthesis of new superheavy elements, has built new facilities and upgraded its existing ones. In the index, the JINR contributes

to 3% of Russia's physical sciences weighted fractional count (WFC), a measure of the relative contribution of an institution to each paper.

Russia's 290-year-old Academy of Sciences (RAS) is the top institution in the region by article count (AC) and WFC. However, RAS is undergoing a painful overhaul. In 2013, president Vladimir Putin's administration passed laws forcing the RAS to merge with its two sister academies, which focus on medical and agricultural research. It is also downsizing, with its 483 institutes facing staff cuts. The turmoil at RAS is not the only blight on the Russian research landscape. Earlier this year, NATO and NASA cut ties with Russia in response to the situation in the Ukraine, threatening the future of its international collaborations, which within the index, at least, are greater than the global average.

REGIONAL REGENERATION

With the collapse of communism in the early 1990s came turmoil, followed by regeneration and change. In central Europe, former Eastern Bloc countries such as Poland, Hungary and the Czech Republic have since been reforming their infrastructures to make them more merit-based.

Poland is second in the region with a WFC of 216 (sandwiched globally between Brazil and Finland). The country, which joined the European Union (EU) in 2004, has been pursuing policies to modernise the traditional,

hierarchical structure of its research institutions. In 2010, new government moves to make the funding of Polish science more competitive included establishing a new funding agency, the National Science Centre, and reforming the existing National Centre for Research and Development. "This is the biggest achievement," says Maciej Żylicz, president of the Foundation for Polish Science (FNP), an NGO that supports scientific research in Poland. About a decade ago, only 13% of government funding was in the form of competitive grants, he says. Today that figure is about 50%. "This is a big change, not only physically, but psychologically."

Poland has also undertaken a number of initiatives to support the life sciences. One notable example is the International Institute of Molecular and Cell Biology (IIMCB) in Warsaw, founded following a 1995 agreement between the Polish government and UNESCO. "This international umbrella helped us to build a structure that is totally different from the structures of the institutes of the Polish Academy of Sciences or universities," says Żylicz, who also heads the Department of Molecular Biology at the IIMCB.

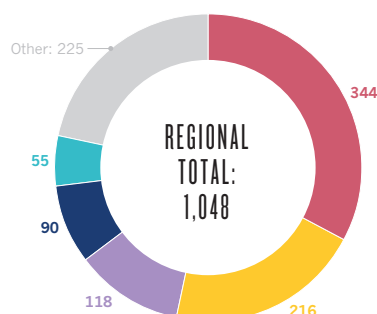
The IIMCB has an international advisory board, which helps select principle investigators. Scientists do not have tenure, but are evaluated every four years and dismissed if they are not up to scratch. "In normal Polish institutes, there is no way to fire professors," says Żylicz.

CENTRAL, EAST & SOUTH EUROPE ANALYSIS

■ Russia ■ Poland ■ Czech Republic ■ Greece ■ Hungary ■ Slovenia ■ Cyprus

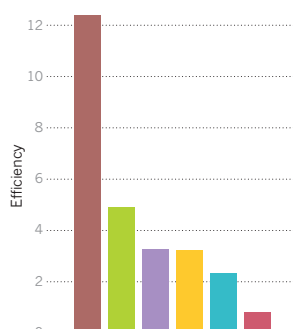
Countries' weighted fractional count (WFC)

Of the 22 countries in this region, Russia and Poland produce more than half of the high-quality science.



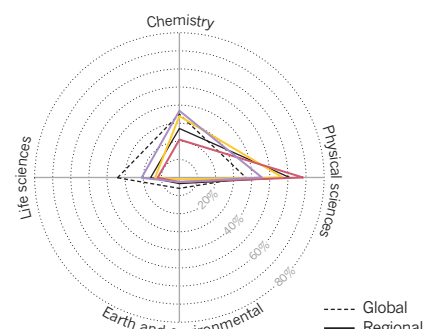
Researcher efficiency

WFC per 1,000 researchers¹.



Research strengths

The whole region is heavily skewed towards the physical sciences².



1. Source: UNESCO. 2. Subjects overlap, so the total for each country can be >100%.

Cyprus may have a population of 866,000 people and a WFC of only 10.9, but it tops the region in terms of its researcher efficiency, or WFC per 1,000 researchers, rivalling that of western European countries such as the United Kingdom and Germany. This might seem surprising given that, unlike these countries, Cyprus' research history is short: its first public university, the University of Cyprus, was only established in 1989; and its first research institute, the Cyprus Institute of Neurology and Genetics, opened in 1990. But, says Vassilis Tsakalos, Director General of the Research Promotion Foundation (RPF), an NGO that supports scientific and technological research, youth is a virtue. "It's a new system," he says. "The institutions were created with the latest ideas in mind. So they are very dynamic."

Although funding in Cyprus suffered as a result of the global financial crisis, scientists are hopeful that spending will now remain stable, says Tsakalos. The RPF gets most of its funding from the government, which spent about half a per cent of GDP on R&D per annum between 2010 and 2012. Recently, the Cypriot government appointed a committee to recommend on the governance of research and innovation.

UNIVERSITY CHALLENGE

Science in this region is dominated by national academies of science, which are conglomerates of tens or hundreds of smaller, dedicated research institutes. To allow for comparison, the index can focus on universities only, which for the top three countries include the University of Warsaw, Charles University in Prague and Lomonosov Moscow State University (MSU). The universities are all of comparable size: Warsaw has around 7,000 staff, and the same number of PhD and postdoc students; Charles University has more than 7,900 staff, including 4,500 academic and research staff; MSU has 10,000 faculty including 5,000 researchers.

Publications from all three are heavily dominated by the physical sciences (see 'Institutional subjects'). Yet while Warsaw is proportionally the most focused on physics, its WFC for this subject

is 21, whereas MSU produces the most physics output in absolute terms with a WFC of 24.

Warsaw is involved in several large international physics collaborations. One to which its researchers contribute particular expertise, says Warsaw's Rector Professor Marcin Palys, is the Compact Muon Solenoid (CMS) collaboration at CERN. Another strength is the university's highly regarded astronomical observatory, which in 2013 made a major contribution to a *Nature* paper that more precisely measured the distance to the Large Magellanic Cloud — an important cosmic yardstick. The paper created a stir internationally, earning the university one of its highest scores on altmetric.com, which tracks media outlets (see 'Online visibility').

Charles University, too, has been successful with international physics projects including the High Energy Stereoscopic System (HESS) Collaboration and the PHENIX Experiment at Brookhaven National Laboratory, as well as projects at CERN. "These collaborations have been very productive," says Jan Konvalinka, vice-rector of Charles University. MSU, likewise, has published extensively as part of the LHCb, CMS and ATLAS collaborations at CERN.

Of the three institutions, Charles University produces the largest fraction (18%) of output in the life sciences. Key to this success are the university's links with the European Molecular Biology Organisation and other EU initiatives, and offers of support to attract foreign scientists.

In terms of publications in *Nature* and *Science*, Charles University lags behind Warsaw and MSU with only 0.3% of its WFC from these journals. Konvalinka expects this to change. As well as investment in facilities, changes in the way the university and government rate papers should increase quality. Previously, the Czech government's science funding schemes valued only the quantity of papers produced, an issue that Konvalinka and several leading Czech scientists have criticized. "Now we need to see to it that not only the quantity of research papers but also the quality of original scientific contributions is increased substantially," says Konvalinka. ■

HUNGARY

Academy of Sciences



The beautiful Hungarian Academy of Sciences

Hungary, like many other countries in central and eastern Europe, has an Academy of Sciences (HAS) as its top institution. But unlike the academies of Russia, the Czech Republic or Poland, for example, HAS has a much stronger record in the life sciences, which comprise more than a third of its output. The academy has a long history in this area: the Biological Research Centre in Szeged and the Institute of Experimental Medicine in Budapest are regarded as leading institutions in the country. The latter contributed to almost half of the academy's WFC in the life sciences.

However, in terms of publications in *Nature* and *Science* HAS is far less productive than the academies of neighbouring countries. It is looking to change this and in 2009 launched its Momentum programme, with Ft2.5 billion (US\$10 million) of government funding, to encourage outstanding young Hungarian researchers to remain in or return to the country. HAS received a further boost in 2012, when the government increased its overall funding by 20% to Ft35 billion (US\$142 million). HAS expects this to translate into a significant increase in publications in the most prestigious journals over the longer term, an academy spokesperson told *Nature*.

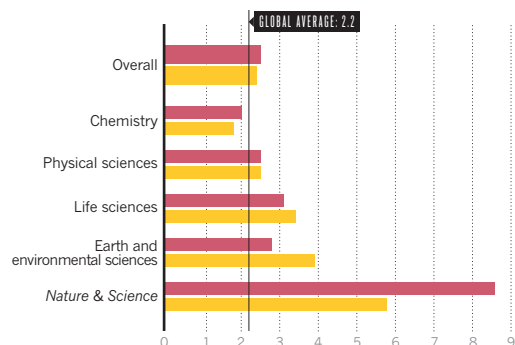
HUNGARIAN ACADEMY OF SCIENCES

Russia Poland

Lomonosov Moscow State University (MSU) University of Warsaw Charles University in Prague

Collaboration rate

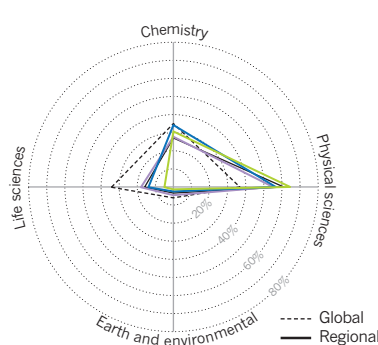
Dividing AC by FC gives a proxy for average number of collaborators per paper from outside the country.



1. Some subjects overlap, so total can be >100%. 2. DOI: 10.1038/nature11878 data from altmetric.com, 22 September 2014.

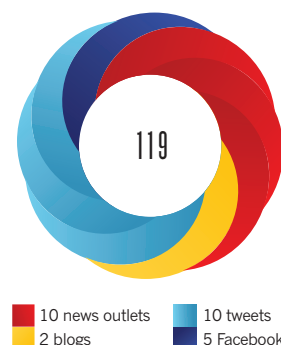
Institutional subjects

Warsaw is most skewed towards physics, but MSU has higher actual output¹.



University of Warsaw's online visibility

A paper involving the astronomical observatory helped refine an important cosmic yardstick².



Australasia & Pacific Islands

The research landscape could not be more disparate between Australia and New Zealand. Strong Australian results are undermined by recent budget cuts, while New Zealand's output has fallen despite science spending boosts.

ARTICLE COUNT: **2,782**

FRACTIONAL COUNT (FC): **1,270**

WEIGHTED FRACTIONAL COUNT (WFC): **1,064**

Australia's weighted fractional count (WFC) of 944 in the Nature Index for 2013 is up from 865 in 2012. Although this keeps it in twelfth place globally, it is one of the few countries that improved its WFC. On the policy front, however, a pall has descended across research this year. When the current conservative coalition government — facing a budget deficit of AU\$41.8 billion (US\$36.5 billion) when it was elected in September 2013 — delivered its first budget in May, there were swingeing cuts across the board in the AU\$9 billion research budget.

Twelve industry innovation programmes, worth AU\$846 million (US\$740 million), were abolished. Another AU\$378 million was lopped from six organizations, including the Australian Research Council, which funds much basic research, and the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the country's largest scientific agency. CSIRO is seventh in the country by WFC in the Nature Index. These cuts meant the loss of several hundred jobs and a decline in output. Environmental and climate science were particularly hard hit — an area of traditional strength for Australia. In the Nature Index, this research accounts for 12% of Australia's output, which is double the global average.

The cuts, along with the government's failure

to appoint a science minister, indicated to many in the science community that research is undervalued. Pre-budget rumours of widespread cuts led Melbourne's *Sunday Age* to warn in an editorial, "this is the era of technology, and to throttle back on funding is more than merely stupid or blinkered; it is vandalism."

**"EXPENDITURE
IN NEW ZEALAND
IS STILL WELL
BELOW THE OECD
AVERAGE."**

The new government did announce the creation of AU\$1 billion dollar Medical Research Future Fund, which it plans to grow to AU\$20 billion by 2020. But this has stalled in the Senate, largely because its funding relies on an unpopular policy to make people pay AU\$7 towards general practitioner visits — the first such fee in the national health scheme.

And more recently, in a statement on 14 October, 2014, the Prime Minister reiterated his view that science matters can be handled by the Minister for Industry. Indeed, the government is attempting to forge closer ties between science and commerce, announcing

an AU\$400 million "Industry, Innovation and Competitiveness Agenda" to fund industry-led non-profit growth centres. Companies may also have a greater role in the school system.

AUSTRALIAN INSTITUTIONS TIED

The index shows that there's little to separate the country's institutions: by article count (AC), the University of Sydney and the Australian National University are the top institutions, having contributed to 474 and 424 papers respectively. WFC gives a measure of the relative contribution of an institution to each paper; using this metric shuffles the order so that the University of Queensland is strongest with a WFC of 95.89, but the University of Sydney (95.07), Monash University (94.69) and the Australian National University (94.09) are close behind.

Focusing on output by subject area reveals further jostling among institutions, with some smaller facilities moving into the top ten in a single subject area, for example Swinburne University of Technology for physical sciences, and the Walter and Eliza Hall Institute of Medical Research in life sciences.

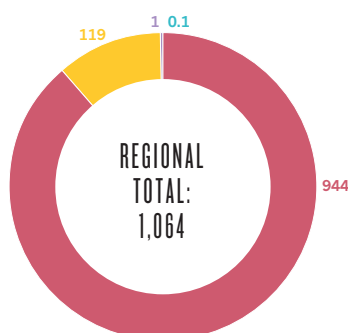
University of Queensland president Peter Høj says the institution's highlights in 2013 included a paper in *Science*, in collaboration with researchers from Massachusetts Institute of Technology, exploring a boson-based

AUSTRALASIA & PACIFIC ISLANDS ANALYSIS

■ Australia ■ New Zealand ■ Papua New Guinea ■ Fiji

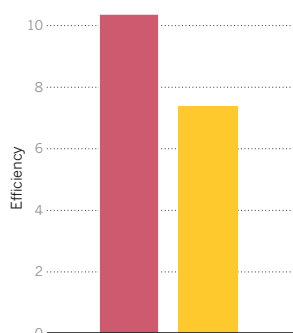
Countries' weighted fractional count (WFC)

Australia accounts for the lion's share of high-quality research.



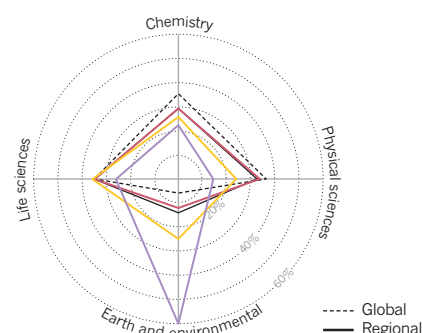
Researcher efficiency

WFC per 1,000 researchers¹.



Research strengths

Smaller countries focus more on earth and environmental sciences².



1. Source: UNESCO (Papua New Guinea and Fiji data not available). 2. Subjects overlap, so the total for each country can be >100%.

alternative to quantum computing, which has potential for the future of quantum devices and supercomputers. This paper was widely shared on social media. Høj adds that such papers illustrate the ability of the university's researchers to participate in "projects with great potential for people worldwide".

Meanwhile, Monash University in Melbourne shows particular strength in the fields of chemistry and life sciences, which heavily overlap. The university is home, for example, to the development of the influenza treatment zanamivir, a neuraminidase inhibitor sold by GlaxoSmithKline under the trade name Relenza. Monash is also strong when it comes to publications in *Nature* and *Science*. A little over 5% of its WFC came from papers in these two journals, which is notably higher than the global average. Its researchers contributed to a total of 15 *Science* or *Nature* papers in 2013, including a report in *Science* on a technique for using graphene as a material for compact energy storage, for which all contributors were from Monash. However, leading the charge on this metric is CSIRO, with 5.5% of its output, or 14 papers, in *Nature* or *Science* (see 'Nature and Science output'). All of these were collaborative works.

NEW ZEALAND BOOSTS ITS BUDGET

Politically, New Zealand is a much different story from Australia. Two days after its neighbour received its bleak science funding news, the New Zealand government revealed a surprise surplus of NZ\$372 million (US\$292 million) in its 2014 budget, and announced an extra NZ\$57 million of science funding in the form of competitive grants as well as NZ\$58 million of increased tax deductions for research and development.

Another NZ\$53 million over four years was allocated to its Centres of Research Excellence programme, and an 8.5% increase – or another NZ\$67.9 million – for university science education.

Nicola Gaston, president of the New Zealand

Association of Scientists, welcomed the boost in funding, but said it was more indicative of the government's shift in focus: the vast majority of competitive grants now require co-funding from industry. Given the nature of the country's industries – largely focused on tourism and primary industries with only a small manufacturing sector, "this has led to serious distortions in the overall balance of science," she said. "Science expenditure in New Zealand is still well below the OECD average, as it has been for some time," she added.

Such a funding deficit is taking its toll. In the Nature Index, New Zealand has a WFC in 2013 of 119, a decrease from its 127 in 2012 (although it retained 28th place).

The University of Auckland contributed to the largest number of papers, with an AC of 123, but on the basis of WFC, the University of Otago topped it. Based in Dunedin and with campuses in Christchurch and the capital Wellington, Otago is the country's oldest university with a good record in the biomedical sciences. Indeed, more than half of the university's WFC comes from the life sciences, with only 13% from the physical sciences (see 'Institutional subject spread').

Richard Blaikie, Otago's deputy vice-chancellor for research, is reluctant to name standouts among the university's 1,100 researchers. But, when pressed, he highlights the university's leadership in an international consortium that identified a new set of genes that instruct stem cells in the human brain how to increase in number and take up position *in utero*. This was one of seven papers the university published in *Nature Genetics*.

The index additionally shows Otago's strength in neuroscience: out of seven papers it contributed to in the *Journal of Neuroscience* in 2013, five were wholly authored by Otago researchers.

Otago's ratio of AC to FC (a proxy for level of collaboration) of 3.0 is above the average for New Zealand of 2.4. Blaikie says this shows that "our staff are engaged in fruitful

PAPUA NEW GUINEA

Pacific positivity

Although Australia and New Zealand dominate the region's science output, the Nature Index also lists papers from the Pacific Island nations. These include one from the Secretariat of the Pacific Community, based in New Caledonia, and five from institutions based in Papua New Guinea, which appears at number 92 on the global list by WFC.

Papua New Guinea's publications in the index focus heavily on environmental science. They include a study from The Nature Conservancy on grouper larvae in coral reefs, and one from Ok Tedi Mining Limited on the effect of the Madden-Julian Oscillation, the largest atmospheric factor in tropical rainfall cycles, on rainfall in the Fly River system in Papua New Guinea. This river system is the wettest place on Earth, and sustained heavy damage from the 1984 OK Tedi Mining disaster – one of the worst environmental disasters ever caused by humans.

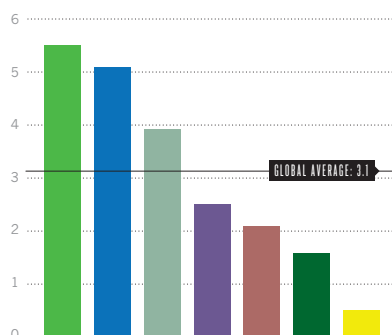
The funding outlook for science in Papua New Guinea is positive, with increases for the Institute of Medical Research and the National Agriculture Research Institute, and a call by the Research, Science & Technology Council for 5% of GDP to be spent on R&D as part of a government-wide National Vision 2050 plan.

international collaborations at the forefront of scientific progress across many areas of enquiry." Some papers also score highly in terms of online visibility as measured by altmetrics.com. Otago researchers were part, for example, of a large global team that published a paper in *Science* detailing how slippery clay was responsible for the 2011 Tohoku earthquake, which was widely picked up by news outlets (see 'Otago's online visibility'). ■



Nature and Science output

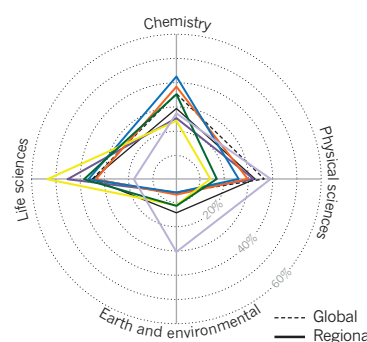
Governmental research agency CSIRO* has the largest proportion of its WFC in these two journals.



*The Commonwealth Scientific and Industrial Research Organisation. 1. Subjects overlap, so the total for each country can be >100%. 2. DOI: 10.1126/science.1238041. Data from altmetrics.com, taken 22 Sept. 2014.

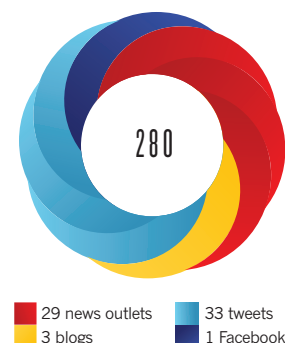
Institutional subject spread

Biomedical sciences are strong for University of Otago, whereas Australian institutions tend to have a more even distribution¹.



Otago's online visibility

Researchers helped reveal how slippery clay was responsible for the 2011 Tohoku earthquake².



West Asia

Almost all countries in the region endured upheaval in 2013. Despite unsettled times the biggest research players continued to prioritize science and discovery and push for breakthroughs across the disciplines.

ARTICLE COUNT (AC): **1,920**
FRACTIONAL COUNT (FC): **785**
WEIGHTED FRACTIONAL COUNT (WFC): **696**

The news from West Asia is often about conflict. Hostilities are played out across borders and within countries. But the picture painted by data in the Nature Index of the regional science landscape shows a dynamic, yet far more collaborative scene.

Israel is the leading science powerhouse in West Asia (see 'Countries' weighted fractional count'). With an article count (AC) over 1,000, its institutions have published more papers in the index than all other regional countries combined. Its research profile is comparable to those of Western countries, for example its publishing record in *Science* and *Nature* is similar to Canada and Australia. It has the highest ratio of researchers to population in the region, and the index reveals that Israel's scientists are amongst the most efficient in terms of weighted fractional count (WFC) per researcher (see 'Researcher efficiency').

In 2012, Israel spent 3.9% of GDP on research and development (R&D), a higher percentage than any other country in the region and among the top five globally. Benjamin Geiger, chair of the Israel Science Foundation (ISF), which supports academic research in Israel, notes that this percentage has declined from a high of 4.2% in 2009. This is not a real decrease in funding but the result of vigorous economic growth: Israel's GDP has risen by 41% since 2009. Manuel Trajtenberg, chair of Israel's planning

and budgeting committee, says this increase is seen in all areas of science spending, "The higher education budget was US\$1.8 billion in 2010, it is now US\$2.6 billion; adjusting for wage inflation, it would be an increase in real terms of 33% in four years." Geiger adds that, during this time, the ISF budget has also increased, from US\$80.2 million to US\$138.3 million.

"WE DO NOT KNOW THE NEXT BIG THING IN SCIENCE, BUT WE CAN KNOW THE OUTSTANDING SCIENTISTS."

Despite its sophisticated output, Israel's scientific growth may have stalled: its WFC in the index is unchanged since 2012, and it has contributed fewer articles in total. Saudi Arabia, by contrast, is a distant second in terms of numbers, but is ramping up its high-quality scientific output. The wealthiest country in the region, Saudi Arabia is increasing its proportional R&D spend too – now approaching 1% of GDP, up from just 0.07% five years ago. And according to the index data, Saudi Arabia is one of the most

efficient countries at getting high-quality science returns for its R&D spend (Overview, page S56).

Some institutes in Saudi Arabia improve their publishing figures by offering attractive packages to overseas researchers, who spend a few weeks or months in the kingdom and add a local affiliation to their papers. This practice shows up in the Nature Index as an unusually high ratio of AC to WFC, because of the higher number of joint affiliations each researcher has. King Abdulaziz University of Science and Technology (KAUST), for example, has an AC of 121 but a WFC of 9.96; other institutions in the index from across the world with the same AC typically have WFCs three-times higher.

KEEPING IT PHYSICAL

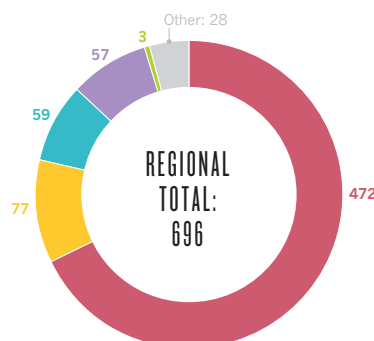
Saudi Arabia's institution with the highest WFC is KAUST, a graduate level university specialising in science and engineering. Located by the Red Sea, KAUST boasts some of the best equipped laboratories and facilities in the region, including one of the fastest supercomputers in West Asia.

KAUST has been actively recruiting top researchers. At its inauguration in 2009, KAUST had 65 faculty hired from different universities around the world, mostly at the assistant professor level; today there are 136. "We are hiring people who are higher calibre.

WEST ASIA ANALYSIS

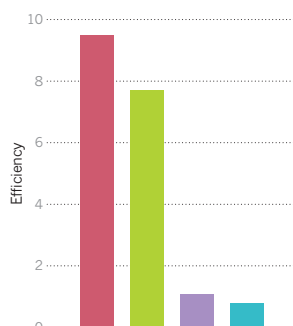
Countries' weighted fractional count (WFC)

Israel accounts for three-quarters of the region's output in the Nature Index.



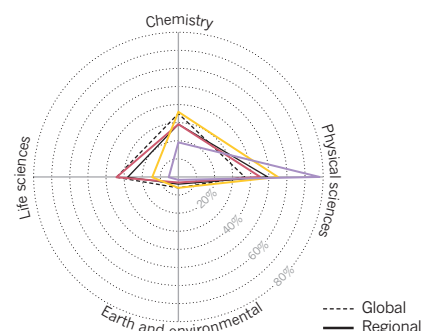
Researcher efficiency

WFC per 1,000 researchers¹.



Research strengths

Emerging scientific countries focus more on physical sciences².



1. Source: UNESCO (Saudi Arabia data not available). 2. Subjects overlap, so the total for each country can be >100%.

We are still in a growth phase and we are very careful not to grow too fast,” says Jean Frechet, vice president for research at KAUST. “We want to grow in quality before quantity, and we want to do it right.”

Frechet himself was recruited from the University of California, Berkeley. When he was approached in 2009, he refused an offer because he had never heard of KAUST.

But he changed his mind after visiting the university, seeing the facilities and meeting other faculty members. “It was very exciting,” he recalls. “Sometimes you get the opportunity to do something that you will not get again.”

The region’s institute with the highest WFC is the Weizmann Institute of Science in Rehovot, Israel.

“SAUDI ARABIA IS CHARACTERIZED BY MATERIAL SCIENCE RESOURCES.”

Notably, it has published 12 papers in *Nature* and *Science*, more than all other institutes in the region, and papers in these journals comprise 4.4% of its output, compared to 0.6% for KAUST. Such a record is the result of progressive hiring policies, says Daniel Zajfman,

president of the Weizmann Institute. “We do not know what the next important thing in science is, but we can know who are the outstanding scientists.”

KAUST and Weizmann have very different research profiles. Life sciences research at KAUST is scant, as it is across Saudi Arabia (see ‘Research strengths’), whereas the Israeli institute has a more even split between chemistry and life and physical sciences. “Saudi Arabia is very much characterized by material science due to its natural resources,” explains Frechet, adding that KAUST modelled itself on the California Institute of Technology (Caltech), which is also heavily focused on physical sciences.

The Weizmann Institute’s science is also much more visible on social media, with five papers (four in the life sciences) receiving a score on altmetrics.com — a new measure of a paper’s impact, determined by article views, downloads and references in social media and news sources — in excess of 100. Its top article was a *Nature* paper authored by three Weizmann researchers collaborating with six researchers from the University of Cambridge and the United Kingdom’s Babraham Institute. This paper challenged previous views about the shape of chromosomes. In contrast, KAUST’s highest paper achieved a score of 47 for describing a new material for scrubbing carbon dioxide from the air. ■

IRAN

Institute for Research in Fundamental Sciences

Located in the Farmanieh district in northern Tehran, the Institute for Research in Fundamental Sciences (IPM) is Iran’s leading institution in the index. It is one of the few top 10 institutes in the region that isn’t in Israel or Saudi Arabia. IPM, credited with bringing the internet to Iran, is almost completely focused on physical sciences research, particularly involving the Large Hadron Collider at CERN (nearly half its papers are in the *Journal of High Energy Physics*).

Seventeen IPM researchers participate in various particle physics experiments, mainly at CERN’s Hadron Calorimeter. Such collaborative work is seen in its high ratio of article count (AC) to weigh fractional count (WFC).

IPM is also trying to engage and stimulate local research through various projects with Iranian universities and research centres. One example is the Iranian Light Source Facility in the northwest Qazvin province, which will house a dedicated synchrotron accelerator once it is complete in late 2018.

Central & South Asia

Asia’s traditional strengths in chemistry and physical sciences continue to power scientific pursuits and collaborations in Central and South Asian countries.

ARTICLE COUNT (AC): **1,574**

FRACTIONAL COUNT (FC): **986**

WEIGHTED FRACTIONAL COUNT (WFC): **879**

Asia’s historic love affair with chemistry and physics is noteworthy. So much so that the topic is a common conversation starter when international scientists meet at a conference dinner.

So it’s not surprising that, in sharp contrast with global publication trends in the Nature Index, 2013 papers from the Central and South Asian region slant steeply toward chemistry and the physical sciences (see ‘Research strengths’). The scientific output of the region is dominated by India — with a population of 1.3 billion and growing — towering over the region’s second-highest performing country and her politically volatile neighbour, Pakistan.

Transitioning from a developing country to an emerging economic superpower, India is

experiencing an attendant surge in its share of the world’s scientific publications. The recovery is largely the result of liberalization, part of the country’s rapid economic growth post-2000. Though allocated funds for science and technology have stagnated at around 1% of GDP over the last two decades, the economic boom means that the absolute amount of money available for scientific research and development has increased. In the 2014 annual budget, India announced a 4% hike in allocations to science-related ministries setting aside 362.69 billion rupees (US\$6 billion) for research.

Nandula Raghuram, a keen metrics watcher and professor at New Delhi-based Guru Govind Singh Indraprastha University, says, “In many consecutive meetings of the Indian Science

Congress, our prime ministers have expressed the need to double the investment for science and technology and bring it to 2% of GDP. It is shameful that this has never happened.”

The Indian government has been urging the private sector to invest more in science but Raghuram says private investment should not be relied on to substitute government funding, which accounts for the lion’s share of science capital in India.

Pakistan, which began its life as an independent country a day before India in August 1947, did not have many scientific institutions when it struck out alone. Despite years of instability and political turmoil it now has a handful of credible scientific institutions. The country spends 0.59% of GDP on science and technology (S&T)

We are still in a growth phase and we are very careful not to grow too fast,” says Jean Frechet, vice president for research at KAUST. “We want to grow in quality before quantity, and we want to do it right.”

Frechet himself was recruited from the University of California, Berkeley. When he was approached in 2009, he refused an offer because he had never heard of KAUST.

But he changed his mind after visiting the university, seeing the facilities and meeting other faculty members. “It was very exciting,” he recalls. “Sometimes you get the opportunity to do something that you will not get again.”

The region’s institute with the highest WFC is the Weizmann Institute of Science in Rehovot, Israel.

“SAUDI ARABIA IS CHARACTERIZED BY MATERIAL SCIENCE RESOURCES.”

Notably, it has published 12 papers in *Nature* and *Science*, more than all other institutes in the region, and papers in these journals comprise 4.4% of its output, compared to 0.6% for KAUST. Such a record is the result of progressive hiring policies, says Daniel Zajfman,

president of the Weizmann Institute. “We do not know what the next important thing in science is, but we can know who are the outstanding scientists.”

KAUST and Weizmann have very different research profiles. Life sciences research at KAUST is scant, as it is across Saudi Arabia (see ‘Research strengths’), whereas the Israeli institute has a more even split between chemistry and life and physical sciences. “Saudi Arabia is very much characterized by material science due to its natural resources,” explains Frechet, adding that KAUST modelled itself on the California Institute of Technology (Caltech), which is also heavily focused on physical sciences.

The Weizmann Institute’s science is also much more visible on social media, with five papers (four in the life sciences) receiving a score on altmetrics.com — a new measure of a paper’s impact, determined by article views, downloads and references in social media and news sources — in excess of 100. Its top article was a *Nature* paper authored by three Weizmann researchers collaborating with six researchers from the University of Cambridge and the United Kingdom’s Babraham Institute. This paper challenged previous views about the shape of chromosomes. In contrast, KAUST’s highest paper achieved a score of 47 for describing a new material for scrubbing carbon dioxide from the air. ■

IRAN

Institute for Research in Fundamental Sciences

Located in the Farmanieh district in northern Tehran, the Institute for Research in Fundamental Sciences (IPM) is Iran’s leading institution in the index. It is one of the few top 10 institutes in the region that isn’t in Israel or Saudi Arabia. IPM, credited with bringing the internet to Iran, is almost completely focused on physical sciences research, particularly involving the Large Hadron Collider at CERN (nearly half its papers are in the *Journal of High Energy Physics*).

Seventeen IPM researchers participate in various particle physics experiments, mainly at CERN’s Hadron Calorimeter. Such collaborative work is seen in its high ratio of article count (AC) to weigh fractional count (WFC).

IPM is also trying to engage and stimulate local research through various projects with Iranian universities and research centres. One example is the Iranian Light Source Facility in the northwest Qazvin province, which will house a dedicated synchrotron accelerator once it is complete in late 2018.

Central & South Asia

Asia’s traditional strengths in chemistry and physical sciences continue to power scientific pursuits and collaborations in Central and South Asian countries.

ARTICLE COUNT (AC): **1,574**

FRACTIONAL COUNT (FC): **986**

WEIGHTED FRACTIONAL COUNT (WFC): **879**

Asia’s historic love affair with chemistry and physics is noteworthy. So much so that the topic is a common conversation starter when international scientists meet at a conference dinner.

So it’s not surprising that, in sharp contrast with global publication trends in the Nature Index, 2013 papers from the Central and South Asian region slant steeply toward chemistry and the physical sciences (see ‘Research strengths’). The scientific output of the region is dominated by India — with a population of 1.3 billion and growing — towering over the region’s second-highest performing country and her politically volatile neighbour, Pakistan.

Transitioning from a developing country to an emerging economic superpower, India is

experiencing an attendant surge in its share of the world’s scientific publications. The recovery is largely the result of liberalization, part of the country’s rapid economic growth post-2000. Though allocated funds for science and technology have stagnated at around 1% of GDP over the last two decades, the economic boom means that the absolute amount of money available for scientific research and development has increased. In the 2014 annual budget, India announced a 4% hike in allocations to science-related ministries setting aside 362.69 billion rupees (US\$6 billion) for research.

Nandula Raghuram, a keen metrics watcher and professor at New Delhi-based Guru Govind Singh Indraprastha University, says, “In many consecutive meetings of the Indian Science

Congress, our prime ministers have expressed the need to double the investment for science and technology and bring it to 2% of GDP. It is shameful that this has never happened.”

The Indian government has been urging the private sector to invest more in science but Raghuram says private investment should not be relied on to substitute government funding, which accounts for the lion’s share of science capital in India.

Pakistan, which began its life as an independent country a day before India in August 1947, did not have many scientific institutions when it struck out alone. Despite years of instability and political turmoil it now has a handful of credible scientific institutions. The country spends 0.59% of GDP on science and technology (S&T)

and is aiming to ramp that up to 2%, by 2020. The country's new science and technology policy tries to connect science with socio-economic development, primarily concentrating on demand-driven research that might help the economy, and through international partnerships. It has some way to go: although India and Pakistan both have roughly the same number of researchers per capita (about 150–160 per million of population), India's scientists are more efficient, producing more than four times as many papers each (see 'Researcher efficiency').

"PAKISTAN'S POLITICS ECLIPSE DEVELOPMENTAL ENDEAVOURS."

Pakistan's political uncertainties continue to eclipse all its developmental endeavours. In a national science and technology policy released in 2012, then science minister Mir Changez Khan Jamali conceded that these exigencies have relegated S&T efforts to the back burner. In the baby steps that Pakistan is taking to shape its research efforts, the focus is on using science to boost the economy through technology transfer projects in metrology, environment, health, energy, biotechnology, agriculture, genetic engineering, electronics and nanotechnology.

Pakistan's Quaid-i-Azam University has an article count (AC) of 52 in the index, the vast majority in the physical sciences (86.1%). The success stems mostly from its natural science faculty, a central part of the university since its foundation. Faculty dean, Mohammed Zakaulah, says most of its research is highly applied.

Two papers in the journal *Applied Physics Letters* that best show the university's strength, wholly authored by its researchers, detail behaviours of "relaxors" – a class of materials that change shape when an electric field is applied – both of which have immediate applications.

SUCCESS STORY

The Indian success story contains highlights of scientific brilliance in recent years, especially in material science, nanosciences and astrophysics, at its many Indian Institutes of Technology (IITs). These technology schools have the highest WFC, which gives a measure of the relative contribution of an institution to each paper, in the region, followed by the government-funded laboratories of the Council of Scientific and Industrial Research and the Indian Institutes of Science Education and Research (IISER). Researchers from the IISERs contributed to three papers in *Nature* and *Science* in 2013, the highest contribution for any Indian institution; the country as a whole only managed 11 papers.

IITs and IISERs are conglomerates or groups of institutes. The standalone institute that shines through is the Indian Institute of Science (IISc) with its formidable chemistry and physical sciences departments producing the highest WFC (83) of the country's individual centres. Institute director Anurag Kumar attributes this success to the fact that new faculty members are provided with start-up research funding, "so that they can get their research programmes off the ground without having to wait for their first grants." IISc also has an ongoing programme that provides seed grants to groups, which most often go on to win large value grants. Of the institution's 132 articles, 60 were wholly authored in-house (an FC of 1), an impressive display of independence. Five of these were in *Physical Review Letters*, including a paper that upped the theoretical mass limit for a star to turn into a type 1a supernova, which did particularly well on Twitter, according to altmetrics.com data.

In the field of physical science, meanwhile, the Tata Institute of Fundamental Research out-produced IISc in 2013, with a WFC second only to the combined IITs. With its stronghold — fundamental research in particle physics and astrophysics — Tata scientists made a mark with their contribution to the CERN experiments that led to the discovery of the Higgs-boson particle. ■

KAZAKHSTAN

L N Gumilyov Eurasian National University



Kazakhstan, a seat of learning in medieval times

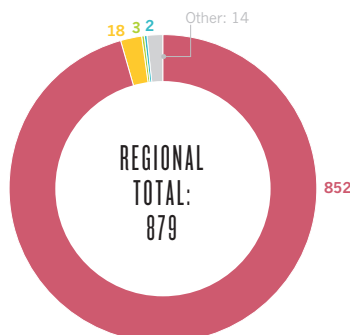
Named after Eurasian scientist Lev Nikolayevich Gumilyov, this national university focuses on the integration of science, education and industry. A former Soviet republic, oil-rich Kazakhstan was considered a major seat of scientific learning in the region in medieval times. Recently, in just a few years, Kazakhstan has more than doubled its science spending from KZT20 billion (US\$109 million) in 2010 to KZT53 billion (\$290 million) in 2013, or around 1% of the country's GDP. The university is also expected to benefit from Kazakh science minister Aslan Sarinzhapov's recent announcement of increasing science spending to 3% of GDP by 2050.

The university might only have six papers in this year's Nature Index, but it is tackling difficult and eclectic areas of physics. Its top paper, authored by three researchers (one with a joint affiliation) from Gumilyov, concerns new exact solutions for static wormholes.

CENTRAL & SOUTH ASIA ANALYSIS

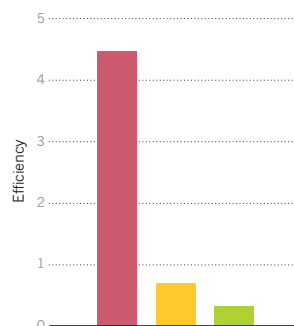
Countries' weighted fractional count (WFC)

Nine countries in this region appear in the index, with India dominating.



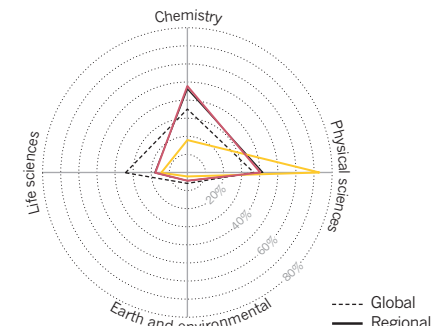
Researcher efficiency

WFC per 1,000 researchers¹.



Research strengths

India and Pakistan have vastly different priorities for their research².



1. Source: UNESCO. 2. Subjects overlap, so the total for each country can be >100%.

Middle & South America

Scientists in Middle and South America are striving for excellence and reaching out for international collaborations, while contending with comparatively low spending levels.

ARTICLE COUNT (AC): **1,968**
FRACTIONAL COUNT (FC): **804**
WEIGHTED FRACTIONAL COUNT (WFC): **530**

Given that more than 92% of the total science and technology investment in Middle and South America goes to Brazil, Argentina and Mexico, it is no surprise that these three countries usually top the regional science output rankings and figure more prominently in the Nature Index than their neighbours. Researchers from this region contributed to 1,968 papers in Nature Index journals in 2013, with a total weighted fractional count (WFC) of 530 — more than three-quarters of which came from these three.

Despite steady growth in the number of peer-reviewed papers published by regional scientists over the past decade, Middle and South America is still far from being a major global player, says Rodolfo Barrere, an Argentinian science metrics specialist who coordinates Ricyt, a survey network compiling data on science policy across all Latin American countries. “The combined effort of all Latin-American countries in the global expenditure in science and technology is between 2% and 3%”, he notes.

Overall, the region’s WFC in Nature Index journals leans towards the physical sciences, boosted by observatories based in Chile, arguably the best country in the world for astronomy.

Argentina has a traditional focus on physical sciences, whereas in Brazil, the discipline

absorbs more attention than it would otherwise because the bureaucracy surrounding life sciences makes it hard to conduct medical research.

At first glance there appears little to distinguish between the three top countries in the region: Brazil’s WFC of 223 places it 23rd globally, Argentina’s 105 puts it at 31st place, and Mexico’s 77 takes the 34th slot (see top 100 table, page S98). All three countries fit the general profile of an emerging nation with

“IN BRAZIL, A SCHOLAR WHO FAILS TO OBTAIN GRANTS STILL HAS HIS WAGE GUARANTEED.”

aspirations of becoming a big league player in science. Yet there are notable differences in terms of how each funds scientific endeavours, and the frequency with which their scientists are published in high-profile journals such as those considered by the Nature Index.

Brazil, where 70% of Latin-American science and technology investments are concentrated, is the only country in the region

where more than 1% of GDP is invested in research and development (R&D). The most recent data compiled by Ricyt, from 2011, showed the country allocating 1.20% of GDP to R&D, followed by Argentina (0.65%), Mexico (0.45%) and Chile (0.44%).

BRAZIL STRIVING TO STIMULATE EXCELLENCE

In Brazil, the University of São Paulo (USP) outstrips other institutions by some margin for article numbers in the Nature Index. It is a huge institution, employing around 6,000 faculty and has around 100,000 graduate and undergraduate students.

USP receives precisely 5.5% of the total commercial tax revenue, (US\$2.2 billion or R\$5 billion) which is the main tributary tool (90%) in São Paulo state. And its researchers are typically awarded half of the US\$500 million in grants offered by Fapesp (São Paulo’s science funding agency).

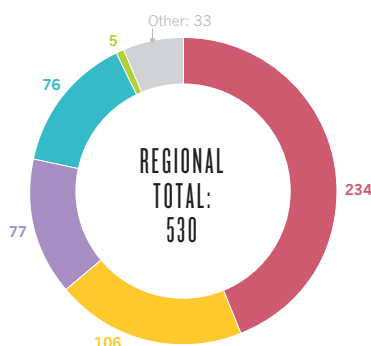
The steady flow of cash in Brazil guarantees resources for scientists, but does not create the most conducive environment to stimulate excellence of quality, according to Jose Eduardo Krieger, deputy dean for research at USP, who compared the system unfavourably to that in the US.

“In the US funding model, the quality assessment of research is embedded in the system,” says Krieger. While a United States researcher’s

MIDDLE & SOUTH AMERICA ANALYSIS

Countries’ weighted fractional count (WFC)

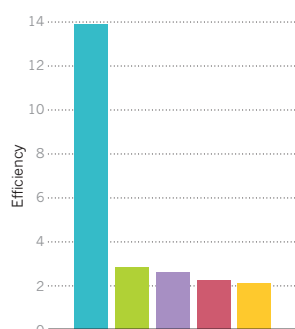
Brazil, Argentina, Mexico and Chile share more than 90% of the regional WFC.



1. Source: UNESCO. 2. Subjects overlap, so the total for each country can be >100%.

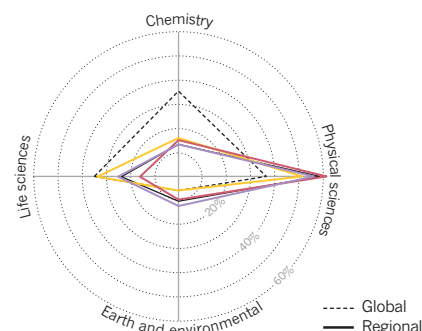
Researcher efficiency

WFC per 1,000 researchers¹.



Research strengths

Only Argentina matches the global average output in life sciences².



income is closely tied to competitive funding, he says, in Brazil, a scholar who fails to obtain grants still has his wage guaranteed.

Within the index, Brazil also has a relatively low rate of international collaboration compared with other Latin-American countries: 2.2 compared to 2.7 for Mexico and 4.3 for Chile.

In order to promote collaboration, the federal government established Science without Borders, a programme under which young academics, mostly undergraduates, go abroad to study.

However Rogério Meneghini, a Brazilian science metrics specialist who directs SciELO, a database of open access journals, questions the value of the project, saying there is no way to be sure these students will continue with a career in science when they return.

RAISING FUNDING TARGETS IN ARGENTINA

The research funding system in Argentina is quite different from Brazil's. The University of Buenos Aires (UBA) is typical of the country's institutions in that it doesn't have a tenure system, explains Alberto Barbieri, the dean of UBA.

He says the work of UBA professors is evaluated every seven years and retaining their chair depends on performance in scientific production, teaching and formation of human resources.

The major challenge in Argentina now

seems to be raising overall funding for R&D. In 2013, the government launched a programme with a target of raising R&D expenditure from 0.65% to 1.65% of GDP, by attracting more private money for science through new public-private projects.

Some researchers are optimistic that this appears to be making Argentina more attractive place to work than flocking to greener pastures abroad. "There are a remarkable number of scientists returning to the country," says Barbieri. Argentina now brags about its academic population, with three researchers per thousand people in the economically active population, twice as many as those doing science in Brazil.

Argentina's leading scientific institution in the index is its National Scientific and Technical Research Council (Conicet), with a WFC of 34. Conicet is a funding body that also has its own labs, and it is these that the Nature Index records.

Conicet and UBA have a similar profile in terms of subject spread, both with a reliance on physical sciences.

However it is third placed National University of La Plata that is the strongest in this field, with nearly four-fifths of research in physics. Part of the reason are the university's two flagship institutes — of Physics and Astrophysics — that are highly collaborative, the former being a part of the ATLAS collaboration at CERN. ■

CHILE

Global connections



The University of Chile stands apart in the region. The university can boast more than 700 international agreements for joint research projects. Many of them are in astronomy, a subject where the university publishes heavily: 94 of its 128 papers are in astrophysics journals.

The University of Chile has a healthy level of collaboration in the index: its ratio of article count to fractional count is 4.8, compared to a global average of 2.2. These collaborations help to leverage the university's 2,500 salaried academics, says Flavio Salazar, deputy rector for research and development. He says the university's scientists are highly productive and their work has significant impact, but are few in number.

Africa

Efforts to boost domestic science spending have yet to reach their goal, but are having an effect in Africa. Researchers, however, still rely heavily on funds and collaborators from richer nations.

Africa is paying attention to science. In 2005 and again in 2014, the continent's leaders met to agree on strategies for investment in science and technology to solve Africa's myriad problems, including ill-health, poverty and political instability. Strategies included funding regional networks in key research areas, such as biosciences and water, and a continental target of countries spending 1% of their GDP on science and technology.

But despite these agreements, no African countries have reached the 1% goal. Today, support for science in Africa varies widely, with countries contributing between less than 0.1% (for Angola) and 0.98% (for Kenya), compared to a world average of 1.7%. Many African scientists still depend partly or wholly on foreign

funding to keep their labs running. As a result, international funders have a strong influence on Africa's research agenda, for example prioritizing advanced vaccine development over low-tech, local health interventions.

The continent's top performing countries in science have long been South Africa followed by Egypt, with Kenya coming in third place. This pattern is reaffirmed by the Nature Index, with South Africa contributing to 441 papers, Egypt 80 and Kenya 24. Strong performances are notable in geology, astronomy and palaeontology; fields in which Africa has geographic advantages such as rich fossil sites or unique geological features.

Patchy funding leads many scientists to collaborate with colleagues from richer continents.

ARTICLE COUNT (AC): 743

FRACTIONAL COUNT (FC): 159

WEIGHTED FRACTIONAL COUNT (WFC): 120

As seen in the ratio of article count (AC) to fractional count (FC), Africa is the most collaborative region (see Global Overview, page S56), and all African countries have a collaboration score well above the global mean of 2.2.

Aside from financial concerns, many Egyptian scientists have difficulty obtaining the chemicals and equipment they need, explains Ramy Aziz, a microbiologist/immunologist at Cairo University. Some collaborate internationally to get what they need; others emigrate and then work with colleagues back home on research projects.

It is slow and costly to import materials and instruments into Egypt, Aziz says — a problem that worsened during the 2011 revolution. It is usual practice to travel to finish off work. For

instance, Aziz spent last summer at the University of California, San Diego, in order to finish six projects that were easier to do there than in Egypt. In the United States, he says, the primers (small pieces of DNA) that he uses for gene amplification can be ordered overnight. In Cairo, it takes up to four weeks to receive them, and costs a lot more.

But, Aziz thinks things are looking up for Egyptian science. The country's government recently increased funds, although in 2012 the science ministry reportedly struggled to spend its enlarged budget due to inefficiencies in the ministry. Egyptian universities and research institutes are also increasingly giving financial incentives for publishing in high-ranking journals. However, change will take time, Aziz says. "Look for the impact of the 2011 revolution in terms of publications by 2020."

BOOSTING RESEARCH QUALITY

The two top-performing African universities in the Nature Index are both South African: the University of the Witwatersrand in Johannesburg and the University of Cape Town (UCT). Both have benefited greatly from two initiatives launched in the last ten years to boost research quality at the country's universities: the South African Research Chairs initiative and the Centres of Excellence scheme. Both have provided quality supervisors for young researchers. To date 15 Centres of Excellence across the country each receive up to R10 million (US\$900,000) per year (a sizeable investment for these institutions), and more than 150 research chairs have been set up with multi-million rand grants attached to them. UCT alone is home to 33 such research chairs.

UCT makes a strong showing in the Nature Index in astronomy and high-energy physics, with more than 70 papers, and does well in many other subjects. Papers in which at least half of the authors are from UCT range from a study contentiously showing that Marine Protected Areas — underwater nature reserves — can improve fishing yields without harming

fisheries, to a report that fledgling birds 'black-mail' their parents into feeding them.

UCT's performance in these fields is not a coincidence, says Marilet Sienaert, executive director of UCT research. Three of the university's government-funded research chairs are in astrophysics, astronomy and marine ecology. It also hosts a government-funded Centre of Excellence looking at birds and biodiversity.

Meanwhile 'Wits', as the University of the Witwatersrand is known, makes a big showing in palaeosciences. It is also head and shoulders over UCT in terms of its publications in *Nature* or *Science*, having 9 articles — accounting for just over 15% of its WFC — in these journals. However, most of these articles come from a special issue of *Science* focusing on *Australopithecus sediba*, an early human ancestor identified from a 2-million year old fossil, discovered by Wits researchers in 2008 at the 'Cradle of Humankind' world heritage site outside Johannesburg.

MANY AFRICAN SCIENTISTS STILL DEPEND PARTLY OR WHOLLY ON FOREIGN FUNDING.

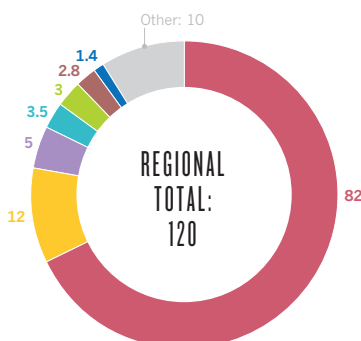
"The special issue on *Australopithecus sediba* has certainly made a significant difference to the count of papers in *Nature* and *Science*," says Robin Drennan, director of research development at Wits. The university was awarded a Centre of Excellence for palaeosciences in 2013. In tandem, the government launched a national palaeosciences strategy providing safe storage facilities for fossils and equipment for casting fossil replicas.

The sediba discovery opened up a new series of caves, all of which show potential for further discoveries, Drennan says. "Our expectation is that a rich and long series of knowledge will flow from this globally unique site." ■

AFRICA ANALYSIS

Countries' weighted fractional count (WFC)

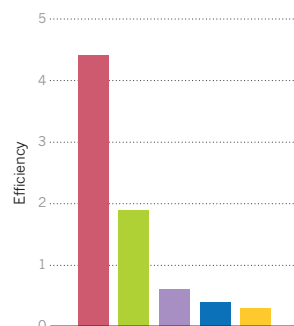
Of the 35 African countries in the index, only 9 have a WFC above 1.



1. Source: UNESCO. 2. Subjects overlap, so the total for each country can be >100%.

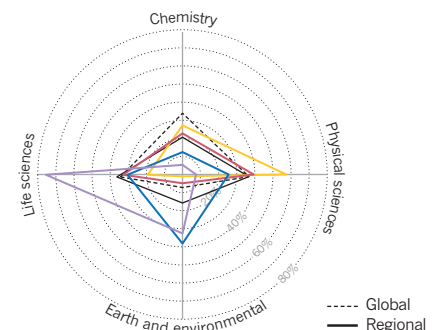
Researcher efficiency

WFC per 1,000 researchers¹.



Research strengths

Africa's natural geology and long human history provide rich sources for the earth and environmental sciences².



HEMIS/LAMY

Addis Ababa University has grand ambitions

Addis Ababa University (AAU) is Ethiopia's largest and oldest research institution. AAU has only eight articles in the Nature Index, but because one was in *Nature* it has a strong ratio in the *Nature* and *Science* measure. The university is strongest in the earth and environmental sciences; five of its articles are on magma upwelling in the country's Afar Depression — the only place on Earth where continental plates are at an advanced stage of separation, giving scientists a first-hand look at how oceans form.

In the Nature Index, AAU is currently 18th overall in Africa. However, it is looking to improve that standing, and has started rewarding staff that bring in external research grants or publish in international journals, says Tassew Woldehanna, vice president for research and technology transfer at AAU. "Our vision is to be one of the top five research universities in Africa," he says.

income is closely tied to competitive funding, he says, in Brazil, a scholar who fails to obtain grants still has his wage guaranteed.

Within the index, Brazil also has a relatively low rate of international collaboration compared with other Latin-American countries: 2.2 compared to 2.7 for Mexico and 4.3 for Chile.

In order to promote collaboration, the federal government established Science without Borders, a programme under which young academics, mostly undergraduates, go abroad to study.

However Rogério Meneghini, a Brazilian science metrics specialist who directs SciELO, a database of open access journals, questions the value of the project, saying there is no way to be sure these students will continue with a career in science when they return.

RAISING FUNDING TARGETS IN ARGENTINA

The research funding system in Argentina is quite different from Brazil's. The University of Buenos Aires (UBA) is typical of the country's institutions in that it doesn't have a tenure system, explains Alberto Barbieri, the dean of UBA.

He says the work of UBA professors is evaluated every seven years and retaining their chair depends on performance in scientific production, teaching and formation of human resources.

The major challenge in Argentina now

seems to be raising overall funding for R&D. In 2013, the government launched a programme with a target of raising R&D expenditure from 0.65% to 1.65% of GDP, by attracting more private money for science through new public-private projects.

Some researchers are optimistic that this appears to be making Argentina more attractive place to work than flocking to greener pastures abroad. "There are a remarkable number of scientists returning to the country," says Barbieri. Argentina now brags about its academic population, with three researchers per thousand people in the economically active population, twice as many as those doing science in Brazil.

Argentina's leading scientific institution in the index is its National Scientific and Technical Research Council (Conicet), with a WFC of 34. Conicet is a funding body that also has its own labs, and it is these that the Nature Index records.

Conicet and UBA have a similar profile in terms of subject spread, both with a reliance on physical sciences.

However it is third placed National University of La Plata that is the strongest in this field, with nearly four-fifths of research in physics. Part of the reason are the university's two flagship institutes — of Physics and Astrophysics — that are highly collaborative, the former being a part of the ATLAS collaboration at CERN. ■

CHILE

Global connections



The University of Chile stands apart in the region. The university can boast more than 700 international agreements for joint research projects. Many of them are in astronomy, a subject where the university publishes heavily: 94 of its 128 papers are in astrophysics journals.

The University of Chile has a healthy level of collaboration in the index: its ratio of article count to fractional count is 4.8, compared to a global average of 2.2. These collaborations help to leverage the university's 2,500 salaried academics, says Flavio Salazar, deputy rector for research and development. He says the university's scientists are highly productive and their work has significant impact, but are few in number.

Africa

Efforts to boost domestic science spending have yet to reach their goal, but are having an effect in Africa. Researchers, however, still rely heavily on funds and collaborators from richer nations.

Africa is paying attention to science. In 2005 and again in 2014, the continent's leaders met to agree on strategies for investment in science and technology to solve Africa's myriad problems, including ill-health, poverty and political instability. Strategies included funding regional networks in key research areas, such as biosciences and water, and a continental target of countries spending 1% of their GDP on science and technology.

But despite these agreements, no African countries have reached the 1% goal. Today, support for science in Africa varies widely, with countries contributing between less than 0.1% (for Angola) and 0.98% (for Kenya), compared to a world average of 1.7%. Many African scientists still depend partly or wholly on foreign

funding to keep their labs running. As a result, international funders have a strong influence on Africa's research agenda, for example prioritizing advanced vaccine development over low-tech, local health interventions.

The continent's top performing countries in science have long been South Africa followed by Egypt, with Kenya coming in third place. This pattern is reaffirmed by the Nature Index, with South Africa contributing to 441 papers, Egypt 80 and Kenya 24. Strong performances are notable in geology, astronomy and palaeontology; fields in which Africa has geographic advantages such as rich fossil sites or unique geological features.

Patchy funding leads many scientists to collaborate with colleagues from richer continents.

ARTICLE COUNT (AC): **743**

FRACTIONAL COUNT (FC): **159**

WEIGHTED FRACTIONAL COUNT (WFC): **120**

As seen in the ratio of article count (AC) to fractional count (FC), Africa is the most collaborative region (Global Overview, page S56), and all African countries have a collaboration score well above the global mean of 2.2.

Aside from financial concerns, many Egyptian scientists have difficulty obtaining the chemicals and equipment they need, explains Ramy Aziz, a microbiologist/immunologist at Cairo University. Some collaborate internationally to get what they need; others emigrate and then work with colleagues back home on research projects.

It is slow and costly to import materials and instruments into Egypt, Aziz says — a problem that worsened during the 2011 revolution. It is usual practice to travel to finish off work. For

instance, Aziz spent last summer at the University of California, San Diego, in order to finish six projects that were easier to do there than in Egypt. In the United States, he says, the primers (small pieces of DNA) that he uses for gene amplification can be ordered overnight. In Cairo, it takes up to four weeks to receive them, and costs a lot more.

But, Aziz thinks things are looking up for Egyptian science. The country's government recently increased funds, although in 2012 the science ministry reportedly struggled to spend its enlarged budget due to inefficiencies in the ministry. Egyptian universities and research institutes are also increasingly giving financial incentives for publishing in high-ranking journals. However, change will take time, Aziz says. "Look for the impact of the 2011 revolution in terms of publications by 2020."

BOOSTING RESEARCH QUALITY

The two top-performing African universities in the Nature Index are both South African: the University of the Witwatersrand in Johannesburg and the University of Cape Town (UCT). Both have benefited greatly from two initiatives launched in the last ten years to boost research quality at the country's universities: the South African Research Chairs initiative and the Centres of Excellence scheme. Both have provided quality supervisors for young researchers. To date 15 Centres of Excellence across the country each receive up to R10 million (US\$900,000) per year (a sizeable investment for these institutions), and more than 150 research chairs have been set up with multi-million rand grants attached to them. UCT alone is home to 33 such research chairs.

UCT makes a strong showing in the Nature Index in astronomy and high-energy physics, with more than 70 papers, and does well in many other subjects. Papers in which at least half of the authors are from UCT range from a study contentiously showing that Marine Protected Areas — underwater nature reserves — can improve fishing yields without harming

fisheries, to a report that fledgling birds 'black-mail' their parents into feeding them.

UCT's performance in these fields is not a coincidence, says Marilet Sienaert, executive director of UCT research. Three of the university's government-funded research chairs are in astrophysics, astronomy and marine ecology. It also hosts a government-funded Centre of Excellence looking at birds and biodiversity.

Meanwhile 'Wits', as the University of the Witwatersrand is known, makes a big showing in palaeosciences. It is also head and shoulders over UCT in terms of its publications in *Nature* or *Science*, having 9 articles — accounting for just over 15% of its WFC — in these journals. However, most of these articles come from a special issue of *Science* focusing on *Australopithecus sediba*, an early human ancestor identified from a 2-million year old fossil, discovered by Wits researchers in 2008 at the 'Cradle of Humankind' world heritage site outside Johannesburg.

MANY AFRICAN SCIENTISTS STILL DEPEND PARTLY OR WHOLLY ON FOREIGN FUNDING.

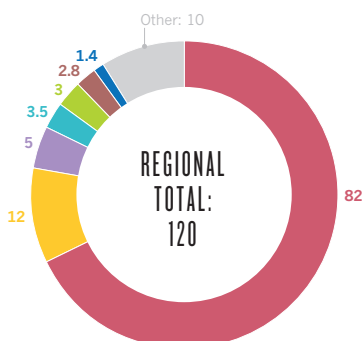
"The special issue on *Australopithecus sediba* has certainly made a significant difference to the count of papers in *Nature* and *Science*," says Robin Drennan, director of research development at Wits. The university was awarded a Centre of Excellence for palaeosciences in 2013. In tandem, the government launched a national palaeosciences strategy providing safe storage facilities for fossils and equipment for casting fossil replicas.

The sediba discovery opened up a new series of caves, all of which show potential for further discoveries, Drennan says. "Our expectation is that a rich and long series of knowledge will flow from this globally unique site." ■

AFRICA ANALYSIS

Countries' weighted fractional count (WFC)

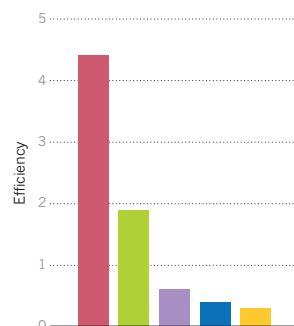
Of the 35 African countries in the index, only 9 have a WFC above 1.



1. Source: UNESCO. 2. Subjects overlap, so the total for each country can be >100%.

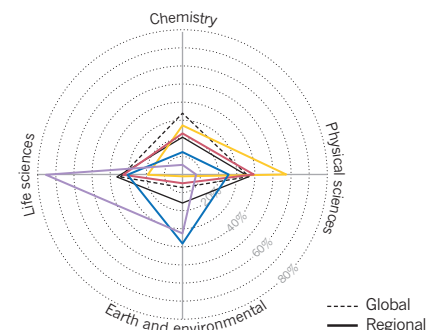
Researcher efficiency

WFC per 1,000 researchers¹.



Research strengths

Africa's natural geology and long human history provide rich sources for the earth and environmental sciences².



HEMIS/LAMY

Addis Ababa University has grand ambitions

Addis Ababa University (AAU) is Ethiopia's largest and oldest research institution. AAU has only eight articles in the Nature Index, but because one was in *Nature* it has a strong ratio in the *Nature* and *Science* measure. The university is strongest in the earth and environmental sciences; five of its articles are on magma upwelling in the country's Afar Depression — the only place on Earth where continental plates are at an advanced stage of separation, giving scientists a first-hand look at how oceans form.

In the Nature Index, AAU is currently 18th overall in Africa. However, it is looking to improve that standing, and has started rewarding staff that bring in external research grants or publish in international journals, says Tassew Woldehanna, vice president for research and technology transfer at AAU. "Our vision is to be one of the top five research universities in Africa," he says.

A guide to the Nature Index

A description of the terminology and methodology used in this supplement, and a guide to the functionality available online at natureindex.com.

The Nature Index is a database of author affiliations and institutional relationships, used to track contributions to articles published in a small group of highly selective journals that have been chosen by an independent group of working scientists.

Data in the Nature Index are updated monthly, with the most recent 12 months of data available under a Creative Commons licence at natureindex.com. The database is compiled by Nature Publishing Group (NPG) in collaboration with sister company Digital Science.

NATURE INDEX METRICS

There are three measures provided by the Nature Index to track affiliation data. The simplest is the article count (AC). A country or institution is given an AC of 1 for each article that has at least one author from that country or institution. This is the case whether an article has one or a hundred authors, and it means that the same article can contribute to the AC of multiple countries or institutions.

To get a better sense of a country or institution's contribution to an article, and to remove the issue of double-counting of articles, the Nature Index uses the fractional count (FC). FC takes into account the relative contribution of each author to an article. The total FC available per paper is 1, and this is shared between all authors under the assumption that each contributed equally. For instance, a paper with 10 authors means that each author receives an FC of 0.1. For authors with joint affiliations, the individual FC is then split equally between each affiliation.

The third measure is the weighted fractional count (WFC), which applies a weighting to the FC in order to adjust for the over-representation of papers from astronomy and astrophysics. The four journals in these disciplines publish about 50% of all papers in international journals in this field — approximately five-times the equivalent figures for other fields. Therefore, although the data for astronomy and astrophysics are compiled in exactly the same way as for all other disciplines, articles from these journals are assigned one-fifth the weight of other articles (i.e. the FC is multiplied by 0.2 to derive the WFC).

Users of natureindex.com can search for specific institutions or countries and generate their own reports, ordered by article count (AC), fractional count (FC) or weighted fractional count (WFC).

Each query will return a profile page that lists the country or institution's recent research outputs, from which it is possible to drill down for more information. For example, articles can be displayed by journal, and then by article title. As in the supplement, research outputs are organized by subject area. The profile page also lists the institution or country's top collaborators, as well as its relationship with other research organizations.

The total FC or WFC for an institution is derived by summing the FC or WFC for individual authors. The process is similar for countries, although complicated by the fact that some institutions have overseas labs that will be counted towards the host country totals. What's more, there is great variability in the way authors present their affiliations. Every effort is made to count affiliations consistently, making reasonable assumptions. For more information on how the affiliation information is processed, please see the frequently asked questions at natureindex.com.

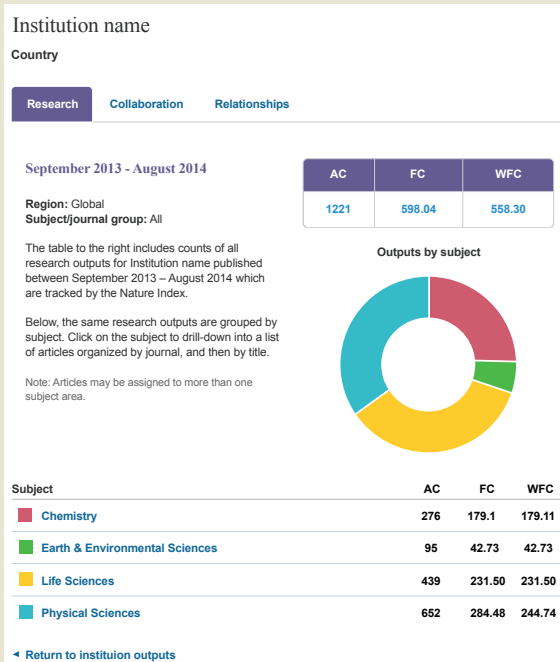
THE SUPPLEMENT

Nature Index 2014 Global is based on a snapshot of data from natureindex.com, covering articles published between 1 January and 31 December, 2013.

Most analyses within the Nature Index 2014

NATUREINDEX.COM

A global indicator of high-quality research



Global supplement use the WFC as the primary metric, as it provides a more even basis for comparison across multiple disciplines, and in determining the relative contribution of each country/institution.

Additional layers of information concerning funding levels, numbers of researchers, size of population and so on, are taken from publicly available sources. In several places, we use altmetrics as a supporting data source. Altmetrics is an alternative way to measure the impact of a paper by tracking different online sources (newspaper stories, tweets, blog posts, comments) that mention the paper. The altmetric score for an article gives an idea of the attention that it has received. Our data are from altmetric.com, provided by the start-up company Altmetric — which is supported by Digital Science. To see more about how this score is calculated, please visit support.altmetric.com. ■

Nature Index tables

The world's leading countries and institutions for high-quality science, ordered by weighted fractional count (WFC) for 2013. Also shown are the total number of articles, and the change in WFC from 2012. Articles are from the 68 natural science journals that comprise the Nature Index (see A Guide to the Nature Index, page S94.)

TOP 100 COUNTRIES

2013	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
1	United States	18,642.88	27,355	18,786.65	-0.8%
2	China	5,205.60	7,637	4,528.97	14.9%
3	Germany	4,076.97	8,669	4,038.30	1.0%
4	Japan	3,370.85	5,102	3,451.26	-2.3%
5	United Kingdom	3,290.35	7,373	3,259.46	0.9%
6	France	2,237.92	5,246	2,343.02	-4.5%
7	Canada	1,483.10	3,220	1,534.98	-3.4%
8	Spain	1,180.25	2,975	1,197.46	-1.4%
9	Switzerland	1,175.18	2,552	1,177.46	-0.2%
10	South Korea	1,150.52	1,953	1,193.02	-3.6%
11	Italy	1,075.12	3,089	1,084.27	-0.8%
12	Australia	943.94	2,448	864.60	9.2%
13	India	851.76	1,380	737.31	15.5%
14	Netherlands	763.24	2,221	765.03	-0.2%
15	Taiwan	543.18	937	595.99	-8.9%
16	Sweden	502.00	1,304	476.73	5.3%
17	Singapore	483.20	833	464.72	4.0%
18	Israel	472.35	1,008	520.05	-9.2%
19	Russia	344.26	1,058	298.26	15.4%
20	Belgium	327.25	1,019	347.22	-5.8%
21	Denmark	298.21	934	301.77	-1.2%
22	Austria	280.61	797	268.85	4.4%
23	Brazil	233.81	670	199.31	17.3%
24	Poland	216.35	689	176.78	22.4%
25	Finland	193.37	586	189.15	2.2%
26	Portugal	124.87	419	114.68	8.9%
27	Norway	123.62	371	142.00	-12.9%
28	New Zealand	118.93	307	127.84	-7.0%
29	Czech Republic	118.43	378	117.83	0.5%
30	Ireland	117.72	336	167.46	-29.7%
31	Argentina	105.66	304	93.08	13.5%
32	Greece	90.15	337	107.55	-16.2%
33	South Africa	81.81	441	58.02	41.0%
34	Mexico	77.08	362	73.52	4.8%
35	Saudi Arabia	76.64	288	52.52	45.9%
36	Chile	75.52	717	71.96	5.0%
37	Iran	58.55	121	71.45	-18.1%
38	Turkey	57.07	202	58.79	-2.9%
39	Hungary	54.92	228	78.88	-30.4%
40	Slovenia	43.20	118	34.85	24.0%
41	Ukraine	28.58	158	35.57	-19.7%
42	Thailand	25.89	89	25.51	1.5%

2013	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
43	Croatia	22.86	105	25.97	-12.0%
44	Pakistan	18.03	63	20.61	-12.5%
45	Romania	17.47	75	21.31	-18.0%
46	Malaysia	14.14	56	5.83	142.7%
47	Serbia	13.95	59	14.84	-6.0%
48	Iceland	13.90	85	15.74	-11.7%
49	Estonia	13.27	68	16.78	-20.9%
50	Slovakia	12.57	65	16.13	-22.0%
51	Egypt	12.04	80	7.32	64.6%
52	Lithuania	11.26	36	7.94	41.7%
53	Cyprus	10.90	25	12.89	-15.5%
54	Luxembourg	10.55	31	5.16	104.5%
55	Colombia	9.62	48	5.84	64.6%
56	Bulgaria	9.49	69	7.30	30.0%
57	Vietnam	8.23	41	6.85	20.1%
58	Armenia	7.98	39	7.30	9.3%
59	United Arab Emirates	7.08	27	2.56	176.7%
60	Panama	6.75	29	8.42	-19.8%
61	Kenya	5.36	24	5.13	4.4%
62	Uruguay	5.21	15	2.93	77.7%
63	Cuba	4.86	16	7.26	-33.0%
64	Indonesia	4.28	22	5.39	-20.6%
65	Peru	4.22	21	3.75	12.5%
66	Qatar	3.96	41	4.01	-1.3%
67	Belarus	3.95	23	6.05	-34.8%
68	Morocco	3.46	21	2.97	16.5%
69	Moldova	3.35	12	3.19	4.8%
70	Lebanon	3.34	17	3.26	2.6%
71	Philippines	3.33	19	2.70	23.5%
72	Latvia	3.18	11	4.04	-21.3%
73	Kuwait	3.17	9	0.43	634.9%
74	Tanzania	3.07	15	0.88	247.5%
75	Kazakhstan	2.85	9	7.20	-60.5%
76	Tunisia	2.84	16	4.52	-37.2%
77	Ecuador	2.48	15	3.29	-24.6%
78	Georgia	2.45	33	3.57	-31.4%
79	Bangladesh	2.28	11	1.23	84.8%
80	Venezuela	1.83	21	4.70	-61.1%
81	Costa Rica	1.49	9	2.11	-29.4%
82	Ethiopia	1.44	10	1.09	32.5%
83	Oman	1.29	6	1.72	-25.1%
84	Tajikistan	1.23	6	0.13	821.6%
85	Greenland	1.22	9	0.12	942.4%
86	Cambodia	1.16	5	0.69	67.4%
87	Algeria	1.10	7	3.51	-68.7%
88	Monaco	1.00	1	1.79	-44.2%
89	Nigeria	0.97	10	0.57	70.7%
90	Barbados	0.92	6	0.00	30144.8%
91	Papua New Guinea	0.91	5	1.04	-12.2%
92	Nepal	0.82	5	1.07	-23.4%
93	Congo	0.82	7	0.12	585.5%
94	Kyrgyzstan	0.81	4	-	-
95	Cameroon	0.81	10	0.60	35.4%
96	Iraq	0.74	9	1.70	-56.5%
97	Botswana	0.71	5	0.38	85.2%
98	Sri Lanka	0.70	7	1.51	-53.6%
99	Mongolia	0.62	7	0.94	-33.4%
100	Jordan	0.60	4	1.58	-61.8%

TOP 200 INSTITUTIONS

2013	INSTITUTION	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
1	Chinese Academy of Sciences (CAS)	China	1,209.46	2,661	1,119.75	8.0%
2	Harvard University	United States	852.12	2,555	904.46	-5.8%
3	Max Planck Society	Germany	728.64	3,023	678.20	7.4%
4	French National Centre for Scientific Research (CNRS)	France	720.62	4,585	718.24	0.3%
5	Stanford University	United States	552.93	1,253	516.78	7.0%
6	Massachusetts Institute of Technology (MIT)	United States	509.49	1,442	486.85	4.7%
7	The University of Tokyo	Japan	474.65	1,293	491.82	-3.5%
8	Helmholtz Association of German Research Centres	Germany	422.47	1,484	393.74	7.3%
9	University of Cambridge	United Kingdom	400.78	1,265	425.47	-5.8%
10	National Institutes of Health (NIH)	United States	390.96	801	386.30	1.2%
11	University of Oxford	United Kingdom	384.03	1,190	339.81	13.0%
12	University of California Berkeley (UC Berkeley)	United States	362.32	1,155	360.74	0.4%
13	University of California, San Diego (UC San Diego)	United States	345.50	873	342.81	0.8%
14	University of Michigan	United States	345.20	905	291.82	18.3%
15	Kyoto University	Japan	313.71	742	312.95	0.2%
16	Northwestern University	United States	298.02	609	294.36	1.2%
17	University of Pennsylvania	United States	296.19	671	241.87	22.5%
18	Yale University	United States	290.60	789	297.67	-2.4%
19	University of California Los Angeles (UCLA)	United States	286.59	768	318.68	-10.1%
20	California Institute of Technology (Caltech)	United States	281.90	1,216	269.89	4.5%
21	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	278.42	847	307.71	-9.5%
22	Peking University (PKU)	China	275.53	743	209.58	31.5%
23	Columbia University in the City of New York	United States	258.74	744	259.25	-0.2%
24	University of Washington	United States	254.74	692	299.79	-15.0%
25	University of Wisconsin-Madison	United States	251.46	652	263.40	-4.5%
26	The University of Texas at Austin (UT Austin)	United States	246.85	595	237.69	3.9%
27	Cornell University	United States	246.17	641	248.14	-0.8%
28	University of Toronto	Canada	242.41	736	262.52	-7.7%
29	Princeton University	United States	232.62	712	223.88	3.9%
30	University of Illinois at Urbana-Champaign	United States	228.61	507	229.51	-0.4%
31	The Johns Hopkins University (JHU)	United States	224.07	753	241.11	-7.1%
32	University of California San Francisco (UCSF)	United States	222.20	474	218.70	1.6%
33	Spanish National Research Council (CSIC)	Spain	220.63	1,553	230.82	-4.4%
34	Osaka University	Japan	219.88	534	261.31	-15.9%
35	Swiss Federal Institute of Technology in Lausanne (EPFL)	Switzerland	207.99	665	186.65	11.4%
36	University of California Santa Barbara (UCSB)	United States	196.36	537	227.84	-13.8%
37	Tsinghua University	China	194.87	474	177.74	9.6%
38	Nanjing University	China	194.57	391	168.10	15.7%
39	National University of Singapore (NUS)	Singapore	191.72	425	186.53	2.8%
40	The University of Chicago (UChicago)	United States	190.55	632	165.48	15.1%
41	Tohoku University	Japan	189.37	455	184.94	2.4%
42	Nanyang Technological University (NTU)	Singapore	187.38	336	178.16	5.2%
43	The Pennsylvania State University (Penn State)	United States	187.20	565	168.05	11.4%
44	Imperial College London	United Kingdom	185.60	668	165.14	12.4%
45	University College London (UCL)	United Kingdom	184.31	738	186.11	-1.0%
46	University of California Davis (UC Davis)	United States	183.99	492	168.74	9.0%
47	University of Minnesota	United States	178.83	449	193.68	-7.7%
48	The Scripps Research Institute (TSRI)	United States	178.38	328	193.05	-7.6%
49	University of Science and Technology of China (USTC)	China	175.73	427	147.75	18.9%

2013	INSTITUTION	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
50	Duke University	United States	174.18	453	162.44	7.2%
51	Washington University in St. Louis (WUSTL)	United States	174.02	376	190.52	-8.7%
52	University of North Carolina at Chapel Hill (UNC)	United States	171.18	373	178.88	-4.3%
53	Russian Academy of Sciences (RAS)	Russia	167.66	879	158.29	5.9%
54	Institute of Physical and Chemical Research (RIKEN)	Japan	165.95	522	144.52	14.8%
55	University of Maryland, College Park (UMCP)	United States	159.00	675	161.42	-1.5%
56	Seoul National University (SNU)	South Korea	158.24	432	178.36	-11.3%
57	Ludwig Maximilian University of Munich (LMU)	Germany	154.14	571	147.63	4.4%
58	Rutgers, The State University of New Jersey	United States	153.88	454	158.52	-2.9%
59	University of Colorado Boulder (CU-Boulder)	United States	152.21	589	147.72	3.0%
60	Lawrence Berkeley National Laboratory (LBNL)	United States	151.89	705	149.60	1.5%
61	Weizmann Institute of Science	Israel	150.96	315	162.04	-6.8%
62	Zhejiang University (ZJU)	China	150.42	289	123.20	22.1%
63	National Research Council (CNR)	Italy	149.79	538	142.41	5.2%
64	University of Pittsburgh	United States	148.19	396	130.26	13.8%
65	Texas A&M University (TAMU)	United States	148.12	414	127.40	16.3%
66	University of California Irvine (UCI)	United States	147.86	452	141.49	4.5%
67	The Ohio State University (OSU)	United States	142.91	553	141.39	1.1%
68	New York University (NYU)	United States	140.07	402	147.54	-5.1%
69	Atomic Energy and Alternative Energies Commission (CEA)	France	138.77	1,088	152.50	-9.0%
70	McGill University	Canada	137.85	459	142.97	-3.6%
71	Georgia Institute of Technology	United States	137.43	293	146.77	-6.4%
72	Nagoya University	Japan	132.96	404	138.68	-4.1%
73	The University of British Columbia (UBC)	Canada	131.61	419	133.51	-1.4%
74	Fudan University	China	129.23	255	121.36	6.5%
75	Indian Institutes of Technology (IITs)	India	128.53	241	125.16	2.7%
76	University of Southern California (USC)	United States	127.44	277	127.69	-0.2%
77	Technical University Munich (TUM)	Germany	125.23	475	130.25	-3.9%
78	The University of Manchester	United Kingdom	124.80	494	117.89	5.9%
79	The University of Texas Southwestern Medical Center at Dallas	United States	124.11	233	126.26	-1.7%
80	University of Copenhagen (UCPH)	Denmark	117.29	547	25.39	361.9%
81	Los Alamos National Laboratory (LANL)	United States	116.13	382	131.73	-11.8%
82	Council of Scientific and Industrial Research (CSIR)	India	115.24	179	107.50	7.2%
83	Nankai University	China	113.77	190	85.85	32.5%
84	Argonne National Laboratory	United States	113.72	450	98.48	15.5%
85	University of Utah	United States	113.04	284	109.78	3.0%
86	Purdue University	United States	112.70	323	141.43	-20.3%
87	Leibniz Association	Germany	112.10	456	120.59	-7.0%
88	University of Bristol	United Kingdom	111.64	378	122.41	-8.8%
89	Utrecht University	Netherlands	109.83	464	109.99	-0.2%
90	The University of Edinburgh	United Kingdom	109.40	641	127.08	-13.9%
91	University of Geneva (UNIGE)	Switzerland	108.05	403	114.75	-5.8%
92	Emory University	United States	107.91	215	102.57	5.2%
93	Hokkaido University	Japan	105.93	240	113.38	-6.6%
94	Tokyo Institute of Technology	Japan	105.72	283	112.64	-6.1%
95	Heidelberg University	Germany	105.61	606	109.59	-3.6%
96	University of Florida (UF)	United States	105.48	415	136.74	-22.9%
97	University of Rochester	United States	104.57	289	87.30	19.8%
98	University of Zurich (UZH)	Switzerland	104.35	426	93.46	11.7%
99	Korea Advanced Institute of Science and Technology (KAIST)	South Korea	103.03	210	126.30	-18.4%
100	Vanderbilt University	United States	101.80	326	136.16	-25.2%

2013	INSTITUTION	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
101	University of Alberta	Canada	101.29	252	99.24	2.1%
102	National Taiwan University (NTU)	Taiwan	101.23	291	96.64	4.8%
103	Boston University	United States	99.30	370	113.82	-12.8%
104	University of Göttingen	Germany	98.99	364	100.98	-2.0%
105	Wuhan University	China	98.80	154	74.27	33.0%
106	Jilin University	China	97.90	179	65.76	48.9%
107	National Aeronautics and Space Administration (NASA)	United States	97.17	978	84.62	14.8%
108	University of Münster (WWU)	Germany	96.69	213	87.80	10.1%
109	Shanghai Jiao Tong University (SJTU)	China	95.99	247	80.03	19.9%
110	The University of Queensland (UQ)	Australia	95.89	308	83.44	14.9%
111	Pierre and Marie Curie University (Paris 6)	France	95.28	1319	91.00	4.7%
112	The University of Sydney	Australia	95.07	474	71.15	33.6%
113	University of Würzburg	Germany	95.02	254	102.53	-7.3%
114	Monash University	Australia	94.69	328	90.99	4.1%
115	Australian National University (ANU)	Australia	94.09	424	85.39	10.2%
116	National Institute of Standards and Technology (NIST)	United States	91.79	286	89.31	2.8%
117	University of Leuven (KU Leuven)	Belgium	91.46	340	98.49	-7.1%
118	The Hebrew University of Jerusalem	Israel	91.29	245	107.24	-14.9%
119	University of Basel	Switzerland	87.67	227	69.70	25.8%
120	Oak Ridge National Laboratory (ORNL)	United States	87.20	283	84.20	3.6%
121	RWTH Aachen University	Germany	87.16	271	83.88	3.9%
122	University of Southampton	United Kingdom	86.68	338	70.04	23.8%
123	University of Notre Dame	United States	86.37	262	88.86	-2.8%
124	The University of Melbourne	Australia	85.79	386	87.15	-1.6%
125	Kyushu University	Japan	84.62	221	107.95	-21.6%
126	University of Erlangen-Nuremberg	Germany	84.58	248	77.00	9.8%
127	Indian Institute of Science (IISC)	India	84.14	132	70.88	18.7%
128	Durham University	United Kingdom	84.05	361	93.87	-10.5%
129	University of Hamburg	Germany	83.31	365	89.25	-6.7%
130	The Rockefeller University	United States	82.62	231	74.49	10.9%
131	Rice University	United States	82.53	227	96.48	-14.5%
132	National Institute for Material Science (NIMS)	Japan	81.79	182	92.60	-11.7%
133	Agency for Science, Technology and Research (A*STAR)	Singapore	81.62	240	86.25	-5.4%
134	Arizona State University	United States	81.20	216	80.87	0.4%
135	Uppsala University	Sweden	80.13	310	87.90	-8.8%
136	Academia Sinica	Taiwan	79.84	394	80.58	-0.9%
137	National Institute of Advanced Industrial Science and Technology (AIST)	Japan	79.53	203	105.78	-24.8%
138	Sun Yat-sen University	China	79.41	158	80.04	-0.8%
139	Iowa State University	United States	78.90	213	87.57	-9.9%
140	Stony Brook University	United States	78.90	284	86.44	-8.7%
141	The University of Arizona (UA)	United States	78.28	550	85.74	-8.7%
142	Aarhus University (AU)	Denmark	78.17	258	82.84	-5.6%
143	University of Bonn (Uni Bonn)	Germany	78.04	426	85.19	-8.4%
144	Tel Aviv University (TAU)	Israel	77.93	289	95.61	-18.5%
145	University of Virginia	United States	77.71	309	82.09	-5.3%
146	University of California Riverside (UCR)	United States	77.58	242	89.77	-13.6%
147	University of Tübingen	Germany	77.39	257	62.15	24.5%
148	Sichuan University	China	76.82	130	44.88	71.2%
149	Case Western Reserve University (CWRU)	United States	76.23	188	91.33	-16.5%
150	Baylor College of Medicine (BCM)	United States	76.10	187	66.26	14.8%

2013	INSTITUTION	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
151	Xiamen University	China	76.02	142	77.84	-2.3%
152	University of Groningen	Netherlands	75.31	375	92.76	-18.8%
153	Memorial Sloan-Kettering Cancer Center (MSKCC)	United States	75.27	184	83.74	-10.1%
154	Pacific Northwest National Laboratory (PNNL)	United States	74.62	178	62.75	18.9%
155	Lund University	Sweden	74.45	252	72.77	2.3%
156	North Carolina State University	United States	74.43	183	51.18	45.4%
157	The University of Nottingham	United Kingdom	74.25	235	77.84	-4.6%
158	Stockholm University	Sweden	73.61	375	57.22	28.7%
159	Leiden University	Netherlands	73.07	509	82.19	-11.1%
160	University of Freiburg	Germany	72.80	242	82.12	-11.4%
161	University of Paris Sud (Paris 11)	France	72.32	604	71.02	1.8%
162	Korea University	South Korea	72.01	220	67.89	6.1%
163	University of Massachusetts Medical School (UMMS)	United States	71.99	149	69.80	3.1%
164	Indiana University Bloomington	United States	71.45	214	68.86	3.8%
165	Brown University	United States	71.43	199	78.32	-8.8%
166	University of London - King's College London	United Kingdom	71.42	214	66.57	7.3%
167	University of Chinese Academy of Sciences (UCAS)	China	71.18	434	64.84	9.8%
168	The University of Hong Kong (HKU)	China	70.43	149	50.45	39.6%
169	The University of New South Wales (UNSW)	Australia	70.14	224	59.18	18.5%
170	The University of Iowa	United States	70.08	258	78.52	-10.8%
171	Pohang University of Science and Technology (POSTECH)	South Korea	70.00	147	84.98	-17.6%
172	Lanzhou University	China	69.99	123	67.58	3.6%
173	Joseph Fourier University (UJF)	France	69.85	596	76.56	-8.8%
174	Delft University of Technology (TU Delft)	Netherlands	69.28	160	64.70	7.1%
175	National Institute for Health and Medical Research (INSERM)	France	68.84	604	71.37	-3.6%
176	University of Barcelona (UB)	Spain	68.73	420	70.73	-2.8%
177	Florida State University (FSU)	United States	68.73	224	80.98	-15.1%
178	University of Calgary	Canada	68.36	147	60.35	13.3%
179	Autonomous University of Madrid (UAM)	Spain	68.16	361	66.86	1.9%
180	Goethe University Frankfurt am Main	Germany	68.14	202	68.20	-0.1%
181	Technion-Israel Institute of Technology	Israel	67.90	194	68.75	-1.2%
182	Brookhaven National Laboratory (BNL)	United States	67.83	268	68.46	-0.9%
183	The University of Sheffield	United Kingdom	67.61	231	70.94	-4.7%
184	University of Glasgow	United Kingdom	67.24	407	66.69	0.8%
185	National Oceanic and Atmospheric Administration (NOAA)	United States	66.86	229	63.25	5.7%
186	Albert Einstein College of Medicine of Yeshiva University	United States	66.10	152	53.18	24.3%
187	The University of Texas MD Anderson Cancer Center	United States	65.99	181	61.19	7.9%
188	East China Normal University (ECNU)	China	65.56	123	35.55	84.4%
189	Soochow University	China	65.30	128	55.27	18.1%
190	National Tsing Hua University	Taiwan	65.16	150	95.15	-31.5%
191	University of Illinois at Chicago	United States	64.44	170	68.23	-5.6%
192	Medical Research Council (MRC)	United Kingdom	64.27	139	60.84	5.6%
193	Radboud University Nijmegen	Netherlands	63.37	438	61.55	3.0%
194	National Chiao Tung University (NCTU)	Taiwan	63.32	124	65.22	-2.9%
195	Michigan State University	United States	62.99	272	73.77	-14.6%
196	University of Montreal	Canada	62.93	233	65.66	-4.2%
197	University of Leeds	United Kingdom	62.81	212	57.71	8.8%
198	IBM Corporation	United States	61.90	117	58.73	5.4%
199	U.S. Navy	United States	61.63	259	69.85	-11.8%
200	Yonsei University	South Korea	61.62	198	77.64	-20.6%

TOP INSTITUTIONS: LIFE SCIENCES

2013	INSTITUTION	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
1	Harvard University	United States	582.14	1,264	609.95	-4.6%
2	National Institutes of Health (NIH)	United States	366.19	755	360.57	1.6%
3	Stanford University	United States	244.26	464	226.04	8.1%
4	Chinese Academy of Sciences (CAS)	China	238.91	478	201.26	18.7%
5	Max Planck Society	Germany	209.85	585	220.13	-4.7%
6	University of California San Francisco (UCSF)	United States	205.95	436	206.55	-0.3%
7	Yale University	United States	190.09	370	178.69	6.4%
8	University of Pennsylvania	United States	187.98	387	165.47	13.6%
9	Massachusetts Institute of Technology (MIT)	United States	186.42	535	161.35	15.5%
10	University of California, San Diego (UC San Diego)	United States	177.45	398	183.78	-3.4%
11	French National Centre for Scientific Research (CNRS)	France	169.30	874	187.51	-9.7%
12	University of Cambridge	United Kingdom	150.50	363	147.77	1.8%
13	University of Michigan	United States	145.39	300	123.65	17.6%
14	The Johns Hopkins University (JHU)	United States	140.95	316	154.27	-8.6%
15	University of Oxford	United Kingdom	139.06	336	146.75	-5.2%
16	University of Washington	United States	138.26	333	161.40	-14.3%
17	Washington University in St. Louis (WUSTL)	United States	134.91	283	150.13	-10.1%
18	Cornell University	United States	134.60	303	117.69	14.4%
19	University of California Berkeley (UC Berkeley)	United States	127.05	268	121.10	4.9%
20	University of Toronto	Canada	126.95	304	145.67	-12.9%
21	The University of Tokyo	Japan	126.93	298	128.57	-1.3%
22	University of California Los Angeles (UCLA)	United States	124.95	285	157.03	-20.4%
23	Columbia University in the City of New York	United States	121.75	282	136.61	-10.9%
24	Duke University	United States	113.63	284	109.88	3.4%
25	University College London (UCL)	United Kingdom	113.22	304	123.73	-8.5%
26	The University of Texas Southwestern Medical Center at Dallas	United States	110.17	214	112.11	-1.7%
27	The Scripps Research Institute (TSRI)	United States	107.52	229	117.37	-8.4%
28	University of California Davis (UC Davis)	United States	96.79	204	76.38	26.7%
29	University of North Carolina at Chapel Hill (UNC)	United States	94.02	217	95.63	-1.7%
30	New York University (NYU)	United States	92.82	223	92.09	0.8%
31	The University of Chicago (UChicago)	United States	92.09	202	92.32	-0.2%
32	University of Wisconsin-Madison	United States	89.69	156	118.77	-24.5%
33	Kyoto University	Japan	88.62	208	76.16	16.4%
34	Northwestern University	United States	88.49	160	90.01	-1.7%
35	Helmholtz Association of German Research Centres	Germany	85.79	408	84.84	1.1%
36	University of Pittsburgh	United States	83.72	171	80.66	3.8%
37	McGill University	Canada	81.63	200	74.81	9.1%
38	Vanderbilt University	United States	74.78	190	93.81	-20.3%
39	Baylor College of Medicine (BCM)	United States	73.74	179	64.70	14.0%
40	Osaka University	Japan	72.73	160	80.62	-9.8%
41	The Rockefeller University	United States	72.41	153	65.70	10.2%
42	Emory University	United States	70.95	152	73.42	-3.4%
43	Institute of Physical and Chemical Research (RIKEN)	Japan	70.63	223	62.12	13.7%
44	University of Minnesota	United States	70.45	148	73.28	-3.9%
45	University of Massachusetts Medical School (UMMS)	United States	68.94	144	67.83	1.6%
46	Rutgers, The State University of New Jersey	United States	67.53	161	66.23	2.0%
47	Spanish National Research Council (CSIC)	Spain	65.91	247	68.06	-3.2%
48	Memorial Sloan-Kettering Cancer Center (MSKCC)	United States	65.88	163	68.89	-4.4%
49	Princeton University	United States	65.71	133	58.69	12.0%
50	University of Rochester	United States	64.85	125	39.37	64.7%

TOP INSTITUTIONS: EARTH AND ENVIRONMENTAL SCIENCES

2013	INSTITUTION	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
1	Chinese Academy of Sciences (CAS)	China	68.50	133	44.78	53.0%
2	Helmholtz Association of German Research Centres	Germany	57.18	143	51.67	10.7%
3	National Oceanic and Atmospheric Administration (NOAA)	United States	56.26	182	53.37	5.4%
4	National Aeronautics and Space Administration (NASA)	United States	51.29	169	43.81	17.1%
5	French National Centre for Scientific Research (CNRS)	France	47.74	295	51.24	-6.8%
6	California Institute of Technology (Caltech)	United States	41.29	116	40.98	0.8%
7	University of Colorado Boulder (CU-Boulder)	United States	37.92	129	39.89	-4.9%
8	University of California, San Diego (UC San Diego)	United States	36.60	85	35.67	2.6%
9	U.S. Geological Survey (USGS)	United States	36.30	99	32.76	10.8%
10	University of Washington	United States	33.87	73	29.14	16.3%
11	The University of Tokyo	Japan	32.72	77	25.19	29.9%
12	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	28.62	82	32.97	-13.2%
13	Columbia University in the City of New York	United States	28.56	81	25.14	13.6%
14	University of Oxford	United Kingdom	27.68	65	18.52	49.5%
15	Stanford University	United States	26.03	57	12.67	105.4%
16	Woods Hole Oceanographic Institution (WHOI)	United States	25.97	61	20.60	26.0%
17	Japan Agency for Marine-Earth Science and Technology (JAMSTEC)	Japan	25.16	70	13.12	91.8%
18	National Center for Atmospheric Research (NCAR)	United States	23.95	82	22.45	6.7%
19	University of Wisconsin-Madison	United States	20.63	50	15.43	33.7%
20	University of Hawai'i at Mānoa (UH Mānoa)	United States	19.83	58	22.67	-12.5%
21	Texas A&M University (TAMU)	United States	19.22	47	13.66	40.6%
22	National Institute of Geophysics and Volcanology (INGV)	Italy	18.53	37	23.64	-21.6%
23	Utrecht University	Netherlands	17.54	46	13.25	32.4%
24	University of California Berkeley (UC Berkeley)	United States	17.47	53	20.22	-13.6%
25	The University of Texas at Austin (UT Austin)	United States	16.71	34	14.33	16.7%
26	Harvard University	United States	16.20	42	14.69	10.3%
27	University of Michigan	United States	16.14	37	13.41	20.3%
28	Colorado State University	United States	15.80	45	13.78	14.7%
29	Australian National University (ANU)	Australia	15.45	41	14.27	8.2%
30	Oregon State University (OSU)	United States	15.30	43	13.69	11.8%
31	Yale University	United States	15.02	28	10.90	37.8%
32	Spanish National Research Council (CSIC)	Spain	14.86	57	10.87	36.7%
33	University of California Santa Barbara (UCSB)	United States	14.52	35	7.90	83.8%
34	University of Cambridge	United Kingdom	14.46	31	23.28	-37.9%
35	The Pennsylvania State University (Penn State)	United States	14.45	40	22.55	-35.9%
36	University of California Davis (UC Davis)	United States	14.33	38	16.80	-14.7%
37	Max Planck Society	Germany	14.31	58	11.49	24.6%
38	The Commonwealth Scientific and Industrial Research Organisation (CSIRO)	Australia	14.28	47	12.32	16.0%
39	University of Florida (UF)	United States	14.06	35	15.50	-9.3%
40	Pacific Northwest National Laboratory (PNNL)	United States	13.84	39	8.29	66.9%
41	Georgia Institute of Technology	United States	13.42	37	13.41	0.1%
42	University of California Los Angeles (UCLA)	United States	13.39	41	11.82	13.3%
43	Institute of Research for Development (IRD)	France	13.33	159	10.96	21.6%
44	Massachusetts Institute of Technology (MIT)	United States	12.99	44	12.61	2.9%
45	Met Office	United Kingdom	12.95	41	20.39	-36.5%
46	University of Maryland, College Park (UMCP)	United States	12.78	46	19.18	-33.4%
47	University of Leeds	United Kingdom	12.66	43	13.98	-9.4%
48	Imperial College London	United Kingdom	12.30	36	6.20	98.4%
49	University of Bristol	United Kingdom	12.25	40	16.63	-26.3%
50	Tohoku University	Japan	12.23	29	9.67	26.5%

TOP INSTITUTIONS: CHEMISTRY

2013	INSTITUTION	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
1	Chinese Academy of Sciences (CAS)	China	679.85	1,148	654.95	3.8%
2	French National Centre for Scientific Research (CNRS)	France	229.20	905	239.62	-4.3%
3	Max Planck Society	Germany	203.77	429	196.43	3.7%
4	Massachusetts Institute of Technology (MIT)	United States	165.70	268	175.91	-5.8%
5	The University of Tokyo	Japan	160.72	263	168.11	-4.4%
6	Kyoto University	Japan	157.00	246	181.97	-13.7%
7	Northwestern University	United States	153.92	236	176.29	-12.7%
8	Stanford University	United States	145.62	226	151.50	-3.9%
9	Harvard University	United States	145.32	249	173.11	-16.1%
10	Peking University (PKU)	China	142.80	268	113.77	25.5%
11	University of California Berkeley (UC Berkeley)	United States	131.34	233	132.02	-0.5%
12	Nanyang Technological University (NTU)	Singapore	123.40	179	127.49	-3.2%
13	University of Oxford	United Kingdom	119.36	197	122.72	-2.7%
14	University of Illinois at Urbana-Champaign	United States	115.67	163	107.03	8.1%
15	Nanjing University	China	115.65	160	97.50	18.6%
16	The University of Texas at Austin (UT Austin)	United States	109.97	151	119.76	-8.2%
17	California Institute of Technology (Caltech)	United States	107.42	145	89.20	20.4%
18	Osaka University	Japan	106.33	187	129.45	-17.9%
19	University of Cambridge	United Kingdom	106.16	182	118.49	-10.4%
20	University of Wisconsin-Madison	United States	104.82	140	110.24	-4.9%
21	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	103.59	197	112.50	-7.9%
22	The Scripps Research Institute (TSRI)	United States	102.50	146	111.01	-7.7%
23	University of Science and Technology of China (USTC)	China	93.78	150	82.17	14.1%
24	University of Michigan	United States	92.58	143	95.67	-3.2%
25	Tsinghua University	China	92.00	169	90.79	1.3%
26	Helmholtz Association of German Research Centres	Germany	91.96	254	74.01	24.2%
27	University of California Los Angeles (UCLA)	United States	87.47	142	106.94	-18.2%
28	Zhejiang University (ZJU)	China	86.42	133	69.32	24.7%
29	Nankai University	China	86.25	124	59.50	44.9%
30	University of Toronto	Canada	84.26	118	94.22	-10.6%
31	National University of Singapore (NUS)	Singapore	83.81	134	70.94	18.1%
32	Swiss Federal Institute of Technology in Lausanne (EPFL)	Switzerland	80.57	150	89.16	-9.6%
33	Fudan University	China	79.94	122	71.59	11.7%
34	University of Pennsylvania	United States	79.89	120	62.05	28.8%
35	University of California, San Diego (UC San Diego)	United States	79.51	141	83.01	-4.2%
36	Council of Scientific and Industrial Research (CSIR)	India	79.46	110	70.56	12.6%
37	University of North Carolina at Chapel Hill (UNC)	United States	75.09	109	83.97	-10.6%
38	Indian Institutes of Technology (IITs)	India	74.77	102	76.14	-1.8%
39	Lawrence Berkeley National Laboratory (LBNL)	United States	74.57	217	67.39	10.7%
40	Jilin University	China	74.06	129	40.87	81.2%
41	University of Minnesota	United States	71.49	98	62.26	14.8%
42	Spanish National Research Council (CSIC)	Spain	70.70	231	77.53	-8.8%
43	Seoul National University (SNU)	South Korea	70.70	121	69.85	1.2%
44	Georgia Institute of Technology	United States	69.30	115	80.59	-14.0%
45	Sichuan University	China	68.45	92	33.30	105.5%
46	University of California Davis (UC Davis)	United States	68.12	111	68.23	-0.2%
47	The Pennsylvania State University (Penn State)	United States	67.92	108	55.20	23.0%
48	Nagoya University	Japan	66.82	101	74.00	-9.7%
49	Texas A&M University (TAMU)	United States	64.05	93	62.87	1.9%
50	University of Washington	United States	63.40	93	77.77	-18.5%

TOP INSTITUTIONS: PHYSICAL SCIENCES

2013	INSTITUTION	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
1	Chinese Academy of Sciences (CAS)	China	412.33	1,234	399.92	3.1%
2	Max Planck Society	Germany	357.86	2,095	322.54	11.0%
3	French National Centre for Scientific Research (CNRS)	France	338.80	2,721	311.35	8.8%
4	Massachusetts Institute of Technology (MIT)	United States	228.44	741	216.24	5.6%
5	Helmholtz Association of German Research Centres	Germany	220.93	778	208.06	6.2%
6	Stanford University	United States	214.72	622	210.97	1.8%
7	The University of Tokyo	Japan	195.25	728	205.31	-4.9%
8	Harvard University	United States	195.03	1,144	223.46	-12.7%
9	University of Cambridge	United Kingdom	169.79	753	184.99	-8.2%
10	University of California Berkeley (UC Berkeley)	United States	139.93	692	151.20	-7.5%
11	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	133.54	496	141.69	-5.7%
12	University of Oxford	United Kingdom	125.31	634	92.75	35.1%
13	University of Michigan	United States	124.60	472	95.49	30.5%
14	Princeton University	United States	118.34	491	124.82	-5.2%
15	University of California Santa Barbara (UCSB)	United States	117.95	376	140.04	-15.8%
16	University of Illinois at Urbana-Champaign	United States	110.64	303	97.94	13.0%
17	Russian Academy of Sciences (RAS)	Russia	110.53	700	106.19	4.1%
18	Peking University (PKU)	China	105.35	388	86.35	22.0%
19	Swiss Federal Institute of Technology in Lausanne (EPFL)	Switzerland	103.24	440	93.83	10.0%
20	California Institute of Technology (Caltech)	United States	102.65	871	121.85	-15.8%
21	University of California Los Angeles (UCLA)	United States	100.93	363	97.65	3.4%
22	The University of Texas at Austin (UT Austin)	United States	100.49	359	99.60	0.9%
23	Tohoku University	Japan	96.67	259	94.84	1.9%
24	National Research Council (CNR)	Italy	93.39	327	89.99	3.8%
25	Atomic Energy and Alternative Energies Commission (CEA)	France	92.22	791	97.41	-5.3%
26	University of Maryland, College Park (UMCP)	United States	91.34	502	89.60	1.9%
27	Northwestern University	United States	90.80	278	88.16	3.0%
28	Kyoto University	Japan	90.46	306	83.76	8.0%
29	Spanish National Research Council (CSIC)	Spain	89.51	1089	100.59	-11.0%
30	Tsinghua University	China	88.12	262	87.16	1.1%
31	National University of Singapore (NUS)	Singapore	87.46	175	91.42	-4.3%
32	Los Alamos National Laboratory (LANL)	United States	84.90	287	96.15	-11.7%
33	Imperial College London	United Kingdom	79.10	365	63.01	25.5%
34	Columbia University in the City of New York	United States	78.16	319	65.23	19.8%
35	University of California, San Diego (UC San Diego)	United States	77.67	300	76.07	2.1%
36	The Pennsylvania State University (Penn State)	United States	77.16	348	64.52	19.6%
37	Lawrence Berkeley National Laboratory (LBNL)	United States	76.89	466	93.79	-18.0%
38	Cornell University	United States	74.62	264	96.13	-22.4%
39	Nanyang Technological University (NTU)	Singapore	72.04	140	69.62	3.5%
40	Georgia Institute of Technology	United States	71.81	155	76.58	-6.2%
41	University of Science and Technology of China (USTC)	China	71.66	242	63.63	12.6%
42	Argonne National Laboratory	United States	71.57	313	62.87	13.8%
43	National Institute of Standards and Technology (NIST)	United States	71.09	235	65.42	8.7%
44	University of Pennsylvania	United States	70.74	220	43.29	63.4%
45	University of Wisconsin-Madison	United States	66.95	351	64.69	3.5%
46	Seoul National University (SNU)	South Korea	66.32	248	75.07	-11.7%
47	Weizmann Institute of Science	Israel	65.64	174	76.86	-14.6%
48	Zhejiang University (ZJU)	China	65.02	136	44.46	46.2%
49	Nanjing University	China	64.94	193	68.24	-4.8%
50	Osaka University	Japan	62.47	222	77.35	-19.2%

TOP INSTITUTIONS IN NATURE AND SCIENCE JOURNALS

2013	INSTITUTION	COUNTRY	WFC	ARTICLE COUNT	2012 WFC	2012-2013 CHANGE IN WFC
1	Harvard University	United States	76.34	199	87.25	-12.5%
2	Massachusetts Institute of Technology (MIT)	United States	45.67	125	35.67	28.0%
3	Stanford University	United States	38.30	87	33.48	14.4%
4	Max Planck Society	Germany	32.86	116	36.51	-10.0%
5	National Institutes of Health (NIH)	United States	27.57	60	22.87	20.6%
6	University of California San Francisco (UCSF)	United States	23.83	45	19.86	20.0%
7	University of California Berkeley (UC Berkeley)	United States	21.54	72	23.13	-6.8%
8	University of Cambridge	United Kingdom	20.65	51	18.92	9.1%
9	Columbia University in the City of New York	United States	19.90	61	9.07	119.4%
10	Chinese Academy of Sciences (CAS)	China	18.64	54	6.26	197.5%
11	The Scripps Research Institute (TSRI)	United States	18.43	38	15.00	22.8%
12	University of California, San Diego (UC San Diego)	United States	17.79	53	23.41	-24.0%
13	Yale University	United States	17.04	43	29.05	-41.4%
14	The University of Tokyo	Japan	15.55	40	10.28	51.3%
15	French National Centre for Scientific Research (CNRS)	France	15.01	117	15.44	-2.8%
16	University of Washington	United States	14.64	38	21.37	-31.5%
17	Swiss Federal Institute of Technology Zurich (ETH Zurich)	Switzerland	13.98	40	18.95	-26.2%
18	Princeton University	United States	13.50	36	13.74	-1.7%
19	University of Toronto	Canada	13.31	34	10.96	21.4%
20	Cornell University	United States	13.11	40	11.26	16.4%
21	University of Oxford	United Kingdom	12.93	45	14.21	-9.0%
22	University of California Los Angeles (UCLA)	United States	12.75	39	13.05	-2.3%
23	California Institute of Technology (Caltech)	United States	12.73	60	17.07	-25.4%
24	The Johns Hopkins University (JHU)	United States	12.33	35	12.05	2.3%
25	The University of Texas Southwestern Medical Center at Dallas	United States	11.79	19	13.20	-10.7%
26	The University of Chicago (UChicago)	United States	11.75	33	11.56	1.7%
27	The Rockefeller University	United States	11.45	23	13.48	-15.0%
28	Helmholtz Association of German Research Centres	Germany	10.75	64	10.75	0.0%
29	New York University (NYU)	United States	10.54	26	10.26	2.8%
30	University of Pennsylvania	United States	10.23	29	7.69	33.0%
31	Institute of Physical and Chemical Research (RIKEN)	Japan	9.62	24	6.80	41.4%
32	University of Colorado Boulder (CU-Boulder)	United States	9.40	35	7.55	24.4%
33	University of Michigan	United States	9.36	40	9.08	3.1%
34	Medical Research Council (MRC)	United Kingdom	9.24	16	7.62	21.3%
35	Swiss Federal Institute of Technology in Lausanne (EPFL)	Switzerland	9.16	26	7.00	30.7%
36	European Molecular Biology Laboratory (EMBL)	Germany	9.05	28	4.02	125.0%
37	National Aeronautics and Space Administration (NASA)	United States	8.77	49	6.84	28.3%
38	Northwestern University	United States	8.61	22	10.24	-15.9%
39	Washington University in St. Louis (WUSTL)	United States	8.41	23	12.74	-34.0%
40	University of North Carolina at Chapel Hill (UNC)	United States	8.38	19	8.42	-0.4%
41	Ludwig Maximilian University of Munich (LMU)	Germany	8.25	24	5.56	48.4%
42	University of California Davis (UC Davis)	United States	8.06	27	5.27	53.0%
43	University College London (UCL)	United Kingdom	7.69	34	7.45	3.2%
44	Kyoto University	Japan	7.48	28	7.91	-5.3%
45	Duke University	United States	7.14	26	11.37	-37.3%
46	The Pennsylvania State University (Penn State)	United States	7.10	20	3.69	92.2%
47	University of Massachusetts Medical School (UMMS)	United States	6.87	14	3.43	100.3%
48	University of Illinois at Urbana-Champaign	United States	6.82	14	7.54	-9.5%
49	Weizmann Institute of Science	Israel	6.67	12	6.45	3.5%
50	University of Minnesota	United States	6.45	19	8.78	-26.5%

Weighted fractional count (WFC) for each institution is shown to two decimal places only. When two or more institutions have the same WFC, their positions are determined by the thousandth place (or beyond).

These results are based on the most recent data available as of 11 September 2014. Owing to continual refinements of the data, the figures in the database are liable to change and might differ to those printed in the supplements.

ADVANCING MATERIALS TO BUILD A BETTER FUTURE

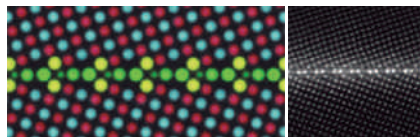


Advanced Institute for Materials Research

The AIMR's top-class international researchers are adding new dimensions to the burgeoning field of materials science, as well as developing innovative functional materials and devices. Their interdisciplinary research is based on atomic and molecular control. The AIMR was established at Tohoku University in 2007 within the framework of the Japanese government's World Premier International Research Center Initiative (WPI) program.

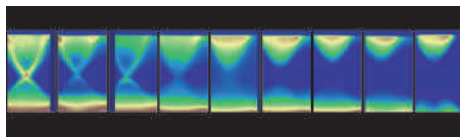


Research Highlights



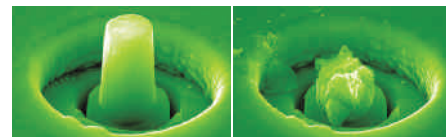
© 2011 Z. Wang *et al.*

Structural defects: Know the boundaries
Z. Wang, *et al.* Atom-resolved imaging of ordered defect superstructures at individual grain boundaries. *Nature* 479, 380 (2011).



© 2011 S. Souma

Insulators: Electrons gaining mass
T. Sato, *et al.* Unexpected mass acquisition of Dirac fermions at the quantum phase transition of a topological insulator. *Nature Physics* 7, 840 (2011).



© 2012 K. M. Reddy *et al.*

Ultra-hard ceramics: Open up and toughen up
K. M. Reddy, *et al.* Enhanced mechanical properties of nanocrystalline boron carbide by nanoporosity and interface phases. *Nature Communications* 3, 1052 (2012).



Web | www.wpi-aimr.tohoku.ac.jp
research.wpi-aimr.tohoku.ac.jp



Flinders University has a wide and active research program addressing some of the critical environmental, health, social, cultural and economic challenges facing our society. Flinders has a number of world class research centres and institutes led by outstanding researchers working across the University and with external academic, government and industry partners. These are some of them.

The Flinders Centre for Innovation in Cancer is the first integrated centre in Australia and one of the few research institutions in Australia with the capability and mix of expertise, to fully address the science of cancer prevention.

The Centre for Neuroscience, established in 1997 is a multidisciplinary collective of researchers and clinicians with a common goal of understanding the brain, spinal cord and peripheral nerves in health and disease.

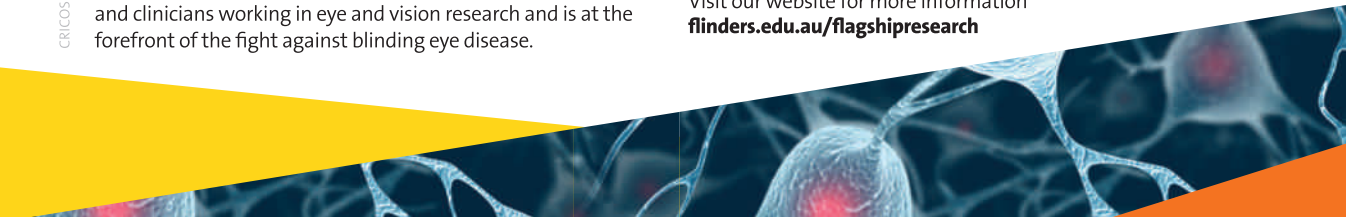
The Southgate Institute for Health, Society and Equity informs practice and policy development about mental and physical health equity in Australia and overseas.

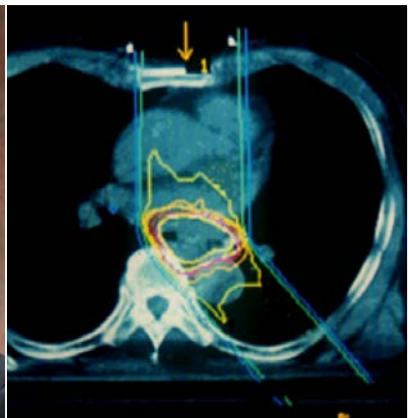
The Flinders Centre for Ophthalmology, Eye and Vision Research brings together one of Australia's largest groups of scientists and clinicians working in eye and vision research and is at the forefront of the fight against blinding eye disease.

The Medical Device Research Institute is a network of over 50 researchers and clinicians who are skilled in the development and application of a diverse range of medical technologies aimed at improving the quality of life for the ageing population both in Australia and across the world. The multi-disciplinary research produces a diverse range of assistive technology from traditional to virtual.

The Centre for Nanoscale Technology undertakes research in partnership with CSIRO, Japan's National Institute for Materials Science and many other esteemed institutions globally. The research areas are in four key areas: Energy, water, security and health, with a goal to make a real difference in these critical areas.

Visit our website for more information
flinders.edu.au/flagshipresearch





GUNMA UNIVERSITY LEADING THE WORLD IN NEXT-GENERATION CANCER THERAPY

Established just four years ago as a state-of-the-art facility for cancer therapy, the Gunma University Heavy Ion Medical Research Center is taking next-generation ion beam therapy even further with precise unrivalled beam control and dose shaping, bringing non-invasive microsurgery closer to reality.

It is an unfortunate fact of modern life that each of us has an almost 50 per cent chance of being diagnosed with cancer within our lifetime. This lifetime cancer risk has risen steadily in developed countries from as low as one in four 40 years ago, and is predicted to continue to rise alongside life expectancy. This close association with advancing age and cancer incidence is of particular importance in Japan.

"Cancer care is a pressing issue for Japan," says Takashi Nakano, director of the Gunma University Heavy Ion Medical Research Center, "not only in terms of developing therapies to increase cancer survival rates, but also in the need for minimally invasive therapies that emphasize a high quality of life."

For many years, X-ray radiation therapy has been one of the few treatment options available for inoperable cancer. Radiation therapy can be effective but the procedure is not sufficiently precise to prevent radiation damage to surrounding healthy tissue. This is particularly true in the brain and for tumours near

radiation-sensitive tissue, as is the case for prostate cancer and pancreatic cancer. In the 1990s, proton beam therapy was introduced in hospitals as an alternative to radiation therapy, offering more precise dose targeting and lower extraneous radiation damage. The technique has since been adopted widely throughout the world with over 40 facilities now available, mostly in Japan, the United States and Europe. However, proton beam therapy still has its limitations.

"Proton beam therapy has superior dose distribution to radiation therapy, but it is not perfectly sharp and still results in some radiation damage to surrounding tissue," explains Nakano. "In addition, because of its similar biological effectiveness to X-rays, some types of tumour tissue are also resistant to proton irradiation, such as hypoxic tumour cells and cancer stem cells, which are recognized as being quite important factors in cancer control. In the 1990s when proton beam therapy had just been introduced, my colleagues and I began researching an alternative beam therapy involving the use of heavy ions, specifically carbon ions, at the National Institute of Radiological Science (NIRS). Twenty years later, we now have this state-of-the-art, next-generation carbon ion radiation therapy centre here in Gunma."

THE MANY BENEFITS OF CARBON IONS

Carbon ion radiation therapy has some very specific advantages as a therapeutic tool even compared to proton beam therapy. For a start, because of its stronger biological effectiveness than X-rays and protons, all cancer cells, even radiation-resistant cells, are affected almost equally by the carbon ion beam, meaning that tumour cells resistant to other forms of radiation therapy, including hypoxic tumour cells and cancer stem cells, are treated effectively. The dose distribution can also be 'shaped' in three



LEARN MORE

Takashi Nakano, Director
Gunma University Heavy Ion Medical
Research Center

Visit: www.heavy-ion.showa.gunma-u.ac.jp/en
Phone: +81-27-220-7111



dimensions to match the exact shape of the tumour.

"We set the shape of the dose distribution by computer using some sophisticated beam focusing equipment," says Nakano. "Carbon ions have much lower scatter than protons, and the dose shape is very sharp, so we can achieve much better dose accumulation in the tumour, and much less damage to surrounding tissue."

"The technology might be expensive and difficult to access now, but in the same way that computers have evolved from room-sized calculators to powerful multi-function devices that can fit in the palm of your hand, it is not hard to imagine that ion beam therapy may one day be found in most cancer centres, providing precise, non-invasive therapies for everything from simple lesions to advanced cancer."

Together, these features of carbon ion radiation therapy make it possible to deliver higher, more effective radiation doses to tumours with almost no damage to surrounding tissue, allowing treatment to be completed in as few as one or two sessions. "We can also treat tumours of any size with equal ease. It means a much higher quality of life for patients, fewer and less severe complications, and we hope much improved survival rates," says Nakano.

Carbon ion radiation therapy offers a real and effective option for inoperable cancer, and for cancer patients with complications such as asthma, which would normally rule out any form of surgery. Although carbon

ion radiation therapy is still relatively expensive due to the limited facilities currently available around the world, the treatment is attractive compared with the other available options, offering the possibility of significantly improved survival rates and much higher quality of life compared with difficult palliative therapies such as chemotherapy.

SURGICAL PRECISION

There are now a handful of heavy ion beam therapy units in operation around the world, but where the Gunma facility stands out is in its unmatched beam control. "We have developed systems that provide control of the position of the beam and the shape of the dose with sub-millimetre precision," explains Nakano. "At this level of precision, highly responsive and accurate respiratory gating is very important, as well as precise knowledge of the location of the tumour. We use computed tomographic image guidance to achieve the sub-millimetre resolution of tumour boundaries we need to target the beam and dose. We are also developing an even more advanced and precise imaging system based on Compton cameras at multiple wavelengths. We are continually pushing the limits of this technology."

With a tool this precise, a range of other uses for the carbon ion beam become possible. One of the most exciting potential uses of the technology is in non-invasive microsurgery. "This is something found nowhere else in the world. We have a treatment room at the Gunma facility set aside for the development of this microsurgery technique. Not only can we treat tumours in very sensitive locations, such as adjacent to the spinal cord, we can also treat other benign diseases, such as vascular lesions, acoustic neuroma near the inner ear, and age-related macular degeneration – all conditions that would be difficult to treat even by a surgeon, yet we expect to be able to achieve even greater surgical precision."

A GLOBAL RESEARCH CENTRE

Nakano has been involved in the development of the carbon ion beam technology

since the idea was first proposed at the NIRS. He then moved to Gunma University as a professor of radiation oncology, where he began work on the carbon ion radiation therapy facility. In collaboration with the NIRS, Nakano's team began developing a compact carbon ion beam technology. With the support of Gunma Prefecture and the technical expertise of Mitsubishi Electric Cooperation, construction of the compact Gunma facility adjacent to Gunma University Hospital began in 2008 and the medical centre was finally opened in 2010. Nakano became head of the new medical centre, and continues to drive its development through the establishment of world-class doctoral programs aimed at cultivating global leaders in ion beam therapy.

"There is a shortage in Japan of the radiation oncologists and medical physicists needed to develop ion beam therapy further," says Nakano. "Our centre has received a significant grant from the Japanese government to run a PhD program to develop the next leaders in this field."

Establishing Gunma Prefecture as a global technological centre of excellence for ion beam therapy is also one of the major aims of the Gunma University Heavy Ion Medical Research Center. Nakano has secure government funding to cultivate the medical and engineering industries needed to support and advance the technology, with the Gunma centre acting as a think tank.

"I believe that in the future, most developed countries will be able to offer carbon ion radiation therapy and enjoy its many benefits," says Nakano. "The technology might be expensive and difficult to access now, but in the same way that computers have evolved from room-sized calculators to powerful multi-function devices that can fit in the palm of your hand, it is not hard to imagine that ion beam therapy may one day be found in most cancer centres, providing precise, non-invasive therapies for everything from simple lesions to advanced cancer."



JAMES COOK
UNIVERSITY
AUSTRALIA

James Cook University:
Highest cited research
institution in the world
for coral reef science.*

CRI/COS Provider Code: 001171

jcu.edu.au

*ISI Essential Science Indicators 2008-2013

Cairns
Singapore
Townsville



KEIO UNIVERSITY IGNITING JAPANESE SOCIAL INNOVATION

Keio university is the oldest and one of the most prestigious universities in Japan, with a history that dates back more than 150 years.

Located in the heart of Tokyo, Keio is a forward-looking, progressive institution founded on principles of scientific realism and Western philosophy, and is focused on enhancing the globalization of education in Japan.

Keio was founded by Yukichi Fukuzawa (1835–1901), an educator and intellectual leader who is considered a pioneer of modern Japan. Fukuzawa was born to a samurai family during the final years of the Edo period. He founded Keio Gijuku in 1858, as a school for Western learning in Edo, today's Tokyo. Keio Gijuku was characterized by its tenets of self-respect and independence of mind combined with an emphasis on *jitsugaku* (science). The school began offering university education in 1890, setting up its first departments in literature, economics and law, and has since continued to establish a firm reputation as a leading university in Japan. Indeed, Keio has a long history of attracting some of the world's brightest minds. For example, in 1922, the year after he won the Nobel Prize in Physics, Albert Einstein visited the university to deliver a five-hour speech, his first in Japan, in front of an audience of over 2,000 people.

Fukuzawa believed the best way for Japan to catch up with Western technology and social systems was to always strive for progress and enlightenment, and provide the academic and moral education needed to create a generation of wise and capable leaders. Graduates of Keio include prominent leaders and historical figures in all walks of life, including former prime minister, Junichiro Koizumi, and highly regarded people in the Japanese business world such as Toyota Motor Corp's incumbent chief executive officer, Akio Toyoda. Notable astronauts, writers, scientists, intellectuals and artists also number among Keio alumni.

This is also reflected by Keio's 9th place ranking in the *Times Higher Education* Alma Mater Index: Global Executives in 2013. The index ranks institutions by the number of degrees they have awarded to CEOs of Fortune Global 500 companies.

Thinking for oneself, taking responsibility for one's actions and *jitsugaku* (science)

Keio University



LEARN MORE

Address: Keio University,
Mita Campus
2-15-45 Mita, Minato-ku,
Tokyo 108-8345

Visit: www.keio.ac.jp



Handwritten manuscripts and letters by Albert Einstein donated to Keio University.



are basic principles held dear by Fukuzawa that Keio continues to uphold to this day. Fukuzawa emphasized the importance of freedom, equality and lifelong learning. He noted, "Heaven does not create one man above or below another man." Fukuzawa is honoured on the 10,000 yen note, Japan's highest denomination.

APPLYING *JITSUGAKU* (SCIENCE) TO SOLVE GLOBAL PROBLEMS

Keio is a comprehensive higher education institution located on six campuses across the greater Tokyo area. It has 10 undergraduate faculties, 14 graduate schools and approximately 30 research centres and institutes, including a university hospital. The institution's affiliates include two primary schools, three junior high schools, and five high schools, including one in New York. Keio is a unique academic entity that offers lifelong education. Keio has partnerships which include 266 overseas institutions and 9 international organizations in 41 countries. Partner universities include Harvard Medical School, Columbia University, Stanford University, the University of Oxford, and the University of London. International organization partnerships include the Asian Development Bank, the International Bank for Reconstruction and Development and UNESCO.

"Keio University continues to make intellectual contributions to Japan and the world to create such affluent ageing societies."

Drawing on its vast research expertise, Keio is now focused on solving challenges in order to realize a prosperous super-mature society. This includes finding ways to promote a sustainable society while

nurturing the next generation and maintaining the health of those in the prime of life and beyond.

"Today, Japan has the world's highest longevity rate. If we can establish our country as a model for a vibrant and prosperous ageing society, it means that we can set a precedent for other countries that are also greying. Keio University continues to make intellectual contributions to Japan and the world to create such affluent ageing societies," says Keio University President Atsushi Seike.

In recognition of its strengths in research and education, Keio University was recently selected by Japan's Ministry of Education, Culture, Sports, Science and Technology as one of the 13 top Japanese universities in the 2014 Top Global University Project. This 10-year grant will further strengthen Keio's capacity to promote research and education in various important fields.

Under the Top Global University Project, Keio will set up three clusters focusing on Longevity, Security, and Creativity, which will all contribute to its goal of attaining sustainable development of a super-mature society. These three multidisciplinary clusters will draw on Keio's unique heritage of innovation and *jitsugaku* (science), and harness the power of the university's strong network of collaborations with industry, academia and international organizations.

For example, the Longevity Cluster will focus on developing innovative solutions to the problems posed by ageing societies. It will adopt a three-pronged holistic approach that will involve considering health matters, sociopolitical issues and technological solutions; examples of each of these aspects include the development of regenerative medicine based on stem cell technology, research into the politics of ageing societies, and the use of robots to undertake nursing duties, respectively.

In the Security Cluster, Keio researchers will explore security issues relating to social, economic and geopolitical risks. Such security issues include global pollution, cyber



Professor Atsushi Seike, president of Keio University.

security in financial markets and regional security in East Asia.

Meanwhile, researchers in the Creativity Cluster will investigate telecommunication and analytical technologies as well as new materials such as plastic optical fibres. In addition, they will investigate new forms of expression in new media and participate in inter-university consortia such as the Global Innovation Design programme. Finally, Keio Business School, the first in Japan to be accredited by the Association to Advance Collegiate Schools of Business, will undertake research on topics that include value creation through outstanding marketing and human resource management.

By consistently emphasizing the philosophy of *jitsugaku* (science), Keio will continue to provide leadership for the future through its commitment to education, research and medicine. Its progressive environment helps students and researchers in their pursuit of excellence, both as researchers and as individuals, and enriches global society.

EXCELLENCE IN RESEARCH

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN



As one of Europe's leading research universities, Ludwig-Maximilians-Universität (LMU) in Munich is committed to the highest standards of excellence in research and teaching. Building on a 500-year-tradition of scholarship, LMU covers a broad spectrum of disciplines, ranging from the Humanities and Cultural Studies through Law, Economics and Social Studies to Medicine and the Natural Sciences. The quality and experience of LMU's faculty and staff form the basis for its distinguished record in research, outstanding performance in the German Excellence Initiative and consistently high rating in national and international rankings.

LMU offers a first-rate academic environment

- A unique and diverse academic culture encompassing all areas of learning and research
- Outstanding research in individual subject areas as well as inter- and transdisciplinary cooperations
- State-of-the-art facilities and conditions for innovative research in all fields of knowledge
- An integral part of the Munich research area – Germany's most vibrant location for science and technology

LMU supports researchers at all career stages

- A structured Academic Career Program for doctoral students and post-docs, enabling them to develop their skills and pursue their research goals in an intellectually stimulating setting
- A well-established and highly successful tenure-track model
- Attractive positions and efficient appointment procedures
- An Equal Opportunities program fostering the talents of female researchers and students





MEIJI UNIVERSITY NEW MODELS FOR ORGAN PRODUCTION AND TRANSPLANTATION

Meiji University was founded in 1881 during the Meiji era in Japan, when the country was going through major political and economic changes, transforming from a feudal society to a modern industrial state. The university was originally named the Meiji Law School, specializing in the instruction of French law to Japanese students with the idea of building a free society founded on individual rights. In 1903, the school was renamed to its current title to reflect its expansion into other areas of study in the humanities as well as the natural sciences. Today, Meiji University has 10 undergraduate schools, 15 graduate schools and 28 departments that cover a comprehensive range of fields, from political science to agriculture, science and technology.

Meiji University has made great strides in research, thanks to the strategic vision of the Organization for the Strategic Coordination of Research and Intellectual Properties established in 2005 to ensure that the university delivers world-class research. The organization consists of two key branches working in close collaboration — the Research Planning and Promotion Headquarters, responsible for the planning, preparation and implementation of university research policy; and the Research Extension and Intellectual Property Headquarters, tasked with

promoting collaborative projects involving industry, academia and the government. Both branches also develop strategies for ensuring a conducive environment for international research at the university.

AGRICULTURE AND MEDICINE

A key institute conducting globally recognized research at Meiji University is the Meiji University International Institute for Bio-Resource Research (MUIBR). Established in 2011, MUIBR is comprised of a network of universities, research institutes and companies in Japan and overseas, and has been awarded several research grants from the Japan Science and Technology Agency, including from the Exploratory Research for Advanced Technology (ERATO) and Core Research for Evolutionary Science and Technology (CREST) funding programmes.

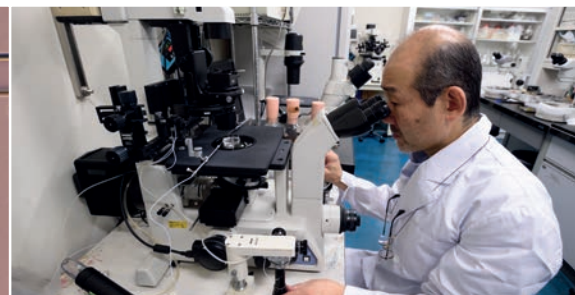
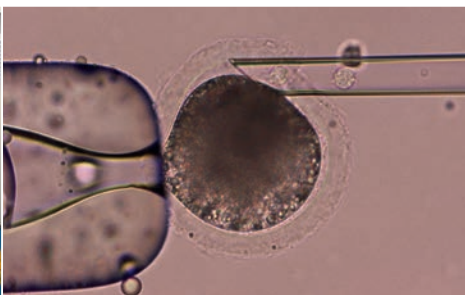
“In agriculture, the concepts of enhancing system efficiency and protecting genetic resources so that they can be passed on to the next generation are very important.”

Hiroshi Nagashima is the director of MUIBR, as well as general manager at the Research Extension and Intellectual Property Headquarters and a professor at Meiji University's School of Agriculture. “In agriculture, the concepts of enhancing system efficiency and protecting genetic resources so that they can be passed on to the next generation are very important,” says Nagashima. MUIBR integrates these agricultural principles and animal



LEARN MORE

Visit: www.meiji.ac.jp/cip/english/
E-mail: osri@mics.meiji.ac.jp



biotechnology methods in the development of new techniques for regenerative medicine and therapies for rare human diseases. To achieve this, MUIBR relies on the creation, maintenance and utilization of biological resources, with a specific focus on genetically engineered pigs.

TRANSLATIONAL RESEARCH

Translational research creates an important bridge between basic biological and medical investigations using model organisms such as mice and clinical research that can contribute to advances in human well-being. The general consensus within the medical community is that pigs are a useful model in preclinical research because of their physiological and anatomical similarities to humans. MUIBR therefore focuses on developing pig models for translational research.

In the area of regenerative medicine, MUIBR uses genetically engineered and cloned pigs to create organs for transplantation. Organ transplantation is the only therapeutic option available for end-stage organ failure, but the procedure is severely limited by shortages in organ availability. The use of induced pluripotent stem (iPS) cells and embryonic stem (ES) cells to generate organs for transplantation offers huge potential for addressing the problem, but difficulties in growing organs derived from iPS and ES cells in cell culture have limited their clinical applicability. Nagashima and his colleagues therefore decided to use pig models to grow organs for transplant surgery.

To fulfil their goal, the researchers first used gene manipulation techniques to create pig models lacking specific organs. They then injected embryonic cells known as blastomeres into the empty space or 'niche' where the original organ would have been to generate an entirely new and viable organ in the vacant niche. In 2010, a team of researchers at the University of Tokyo succeeded in generating a functional pancreas derived entirely from rat iPS cells

in a mouse model. Following this, MUIBR succeeded in taking cells derived from one pig to grow a pancreas in another pig that originally lacked the organ. Researchers at the institute plan to continue these *in vivo* experiments in organ production using different cross-species combinations of model organisms and cells derived from pigs, sheep, cows and even humans.



Hiroshi Nagashima, director of the Meiji University International Institute for Bio-Resource Research, is leading research at the confluence of agriculture and medicine.

In addition to cross-species organ growth, researchers at MUIBR are also investigating "xenotransplantation", in which pig organs are transplanted to humans. The biggest challenge in applying this method is the potential for the recipient's immune system to reject the foreign organ. One form of rejection known as hyperacute rejection can occur within minutes of transplantation. A key factor in activating hyperacute rejection in pig-to-human transplantation is the presence of natural antibodies in humans against galactose epitopes, which are synthesized on the surfaces of pig cells by the enzyme $\alpha 1,3$ -galactosyltransferase. Researchers at MUIBR therefore developed pig models in which the gene encoding $\alpha 1,3$ -galactosyltransferase expression was removed, or knocked out. Only a few researchers in the world have been able to develop

$\alpha 1,3$ -galactosyltransferase-deficient pig models, and the MUIBR team was the first to do so in Japan.

NEW MODELS FOR HUMAN DISEASE

Another major project at MUIBR is the development of pig models for human disease. While many researchers have created mouse models for human disease, the observable characteristics of a mouse model can differ from those phenotypes observed in humans with the same disease. Pigs, on the other hand, are often characterized as having symptoms that more closely resemble those observed in humans. To date, MUIBR has created pig models for Marfan syndrome, a genetic disorder that affects the connective tissue in the body. The gene expressing for the disorder is dominant, meaning that it is expressed even if only one copy is inherited. Approximately 1 in 5,000–10,000 individuals worldwide have inherited the disease.

Ultimately, the clinical application of MUIBR's research will depend on the type of organ or disease targeted for research. One prime candidate for clinical application is the transplantation of insulin-producing cells in the pancreas known as the islets of Langerhans. These islet cells deteriorate as soon as the pancreas is damaged. Given the shortage of organ donors, the techniques of growing organs in animal models and transplanting animal organs into patients could offer a temporary solution until clinically applicable methods of pancreas organogenesis from patient-derived iPS cells are developed.

MUIBR's holistic approach to xenoregeneration-based organ transplantation therapy — from bio-resource development to organ growth and eventual transplantation — has defined its success in the field. Of particular advantage has been the advanced cryopreservation technology established at the institute for safely storing frozen pancreatic islet cells and fertilized eggs. "We are probably the only institute in the world taking such an integrated approach," says Nagashima.



NANYANG TECHNOLOGICAL UNIVERSITY RISING UP THE RANKS

Nanyang Technological University (NTU Singapore) is a young, research-intensive university that is rising in global rankings. Ranked first among top Asian universities in terms of the Normalised Citation Impact (Thomson Reuters InCites 2014) and also in field-weighted citation impact (Elsevier SciVal 2014), NTU has firmly established itself as a powerhouse of research and innovation. It also made a significant leap from 35th place in 2012 to 12th in 2013 in the *Nature Publishing Index Asia-Pacific* ratings, which are based on the number of primary research articles published in Nature-branded journals. This impressive performance is largely due to NTU attracting both ambitious young researchers and established, high-profile scientists.

HOT RESEARCH

Among the top names at NTU Singapore is Professor Zhang Hua (pictured above second from right), an expert in synthesising and fabricating nanomaterials and carbon sheets such as graphene for use in biosensing, clean energy, water treatment and a host of other applications. With a consistent record of highly cited research over the last decade, Zhang received special mention as one of the world's top 17 scientists in the Thomson Reuters 2014 report on *The World's Most Influential Scientific Minds 2014*. "These individuals," the report states, "are influencing the future direction of their fields, and of the world." In 2013 alone, Zhang authored 16 of the highest cited papers in the field of materials science.

Zhang is one in a long list of superstar scientists at NTU. Led by world-renowned

scientist and NTU President Bertil Andersson, winner of the prestigious Wilhelm Exner Medal, NTU's stellar faculty includes NTU Provost Freddy Boey, inventor of innovative biomedical devices; geneticist Stephan Schuster, who co-led the Mammoth Genome Project to sequence the genome of an extinct animal for the first time; Nikolay Zheludev, nanophotonics and metamaterials expert; structural biologist Daniela Rhodes FRS (pictured above far left), distinguished for her research on the structure of chromosomes; geologist and seismologist Kerry Sieh, who has studied earthquakes in the United States, Taiwan and Myanmar, and is addressing the possibility of another devastating quake in Indonesia. Since 2008, Sieh has headed the Earth Observatory of Singapore (EOS), one of two Research Centres of Excellence at NTU.

MANY FACETS OF SCIENCE

Singapore's National Research Foundation and Ministry of Education established the Research Centres of Excellence programme in 2007 to attract leading researchers, boost graduate education and generate new knowledge in specific fields. Fulfilling this mission, the EOS is conducting fundamental research on the interrelated challenges of climate, governance and natural hazards associated with earthquakes and volcanoes. One more national Research Centre of Excellence based at NTU, the Singapore Centre on Environmental Life Sciences Engineering, is researching microbial biofilms with the goal of improving environmental and public health.

These centres draw on NTU's strengths in interdisciplinary research, with specific

capabilities in sustainability, engineering and the life sciences. NTU is ranked 9th in the world for Engineering and Technology (QS World University Rankings 2014), and since 2005 the university has won SGD 1.3 billion in funding to address many facets of global sustainability, including water, energy, climate and public health.

NTU Singapore partnered with Imperial College London to set up the Lee Kong Chian School of Medicine, which welcomed its first cohort in 2013. This was Imperial's first overseas foray. An important pillar of NTU's life sciences cluster, the NTU medical school received SGD 400 million in philanthropic support at its inception. It complements the university's strengths in environmental life sciences engineering, biological sciences and structural biology. Scientists affiliated with the school's laboratories are studying the complex factors associated with metabolic, infectious, neurological and skin-related diseases.

In addition, strong links with companies such as BMW, Johnson Matthey, Rolls-Royce and Lockheed Martin ensure that NTU's excellence in basic and applied research translates into highly relevant industry applications. With its well-established track record in working with the industry, NTU was ranked joint No. 1 globally in Industry Income: Innovation by Times Higher Education in the last two years.



**NANYANG
TECHNOLOGICAL
UNIVERSITY**

LEARN MORE

Visit: www.ntu.edu.sg
[www.twitter.com/NTUsg](https://twitter.com/NTUsg)
www.linkedin.com/company/ntusg
www.facebook.com/NTUsg
 Phone: +65 6791 1744



OKAYAMA UNIVERSITY FIGHTING VIRAL INFECTIONS FOR AGRICULTURE AND MEDICINE

Okayama University was established as a national comprehensive university in 1949 and is located in western Japan, midway between Osaka and Hiroshima on Japan's main island. With 1,600 faculty and more than 13,000 students, the university has grown considerably from a small institute called the Medical Training Place founded in 1870.

The university has a long history of conducting research in the areas of plant and medical science. In 2013 and 2014, it won several highly competitive government grants, including assistance under three high-profile programmes of the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT), namely the Program for Promoting the Enhancement of Research Universities, the Translational Research Network Program and the Top Global University Project. This funding will support more international travel among students and researchers and enhance the international profile of the university by improving its research development.

STRONG BINDING

Professor Takashi Sera from Okayama University's Department of Bioscience and Biotechnology is building on the university's dynamic history through his work on fighting viral infections. As the recent Ebola outbreak has so shockingly demonstrated, viruses can have a devastating impact on human health. They also cause widespread damage to crops and livestock. What makes them so difficult to combat, is the fact that they are moving targets — an antiviral

drug can be developed for a specific virus, but the virus may later mutate into a form that is resistant to that drug. Thus, the war against viruses becomes an endless cat-and-mouse game.

"If we could make cassava, one of the main crops in Africa, resistant to viral infection, we could save people living in Africa from food crises."

Sera is optimistic about a strategy that has the potential to break this cycle. His new approach seeks to prevent viruses from proliferating after an infection occurs. It is an ingenious strategy involving the creation of an artificial DNA-binding protein that attaches to a virus more strongly than the virus's own replication protein.

Sera hopes to use these artificial DNA-binding proteins to produce agricultural crops with greater resistance to viruses and to develop effective antiviral drugs that are side-effect free. He has demonstrated the effectiveness of this technique in both plants and animals. In particular, he has used it with success against the beet severe curly top virus, which is spread through the beet leafhopper and affects hundreds of plant species primarily in the western United States.

NO SIDE EFFECTS

Sera has also used the power of this technique to fight viruses that afflict humans. Specifically,

he has developed antiviral drugs for human papillomavirus, which is implicated in cervical cancer. While vaccines for cervical cancer have already been developed, they have subsequently been reported to have various side effects. Sera is confident that his antiviral drugs will be side-effect free, as they are made of only slightly modified versions of proteins that are naturally found in meat and many vegetables.

The potential of this method is huge, he says. Its benefits could include almost total prevention of viral proliferation with a low probability of creating resistant viruses. Looking to the future, Sera is excited about the possibilities. "If we could make cassava, one of the main crops in Africa, resistant to viral infection, we could save people living in Africa from food crises," he explains. "And if we could create drugs for bird flu or foot-and-mouth disease, we could prevent these infections from spreading to livestock in surrounding areas." The economic benefits of such breakthroughs would be enormous.

"Conducting research at Okayama University has been important for me," Sera says. "The university's long history of promoting a culture of interdisciplinary research as well as its excellent facilities have been vital." As Sera's research demonstrates, this combination of culture and facilities promises to yield great results in the future.



岡山大学
OKAYAMA UNIV.

LEARN MORE

Address: 1-1-1, Tsushima-Naka, Kita-ku, Okayama, 700-8530, Japan

Visit: www.okayama-u.jp/index_e.html
E-mail: ura-info@okayama-u.ac.jp



SUN YAT-SEN UNIVERSITY CHALLENGING TRADITIONAL PARADIGMS

Sun Yat-sen University was founded in 1924 by Dr Sun Yat-sen, a great democratic revolutionary leader of the twentieth century. Its medical school was established even earlier in 1866. The four campuses of Sun Yat-sen University are situated in Guangzhou and Zhuhai, China by the Pearl River and the South China Sea. It is a leading, internationally renowned comprehensive university in China, which advocates a dedication to learning, seeks to imbue people with a passion for education and adopts the mentality that “the professors are the university”. It has one of the most relaxed and unconstrained intellectual atmospheres of all the universities in China.

Sun Yat-sen University is active in a wide variety of academic disciplines, including the liberal arts, social sciences, sciences, engineering and medicine. According to the Essential Science Indicators, 16 academic disciplines at Sun Yat-sen University are ranked in the top 1 per cent worldwide; this number is the third highest for universities in China.

The university boasts eight affiliated hospitals that together treat more than 10 million patients each year. This represents a sophisticated network of medical services and provides a source of abundant first-hand clinical data. The academic strength, large size and high quality of the medical school contribute to it being ranked among the top universities in the country.

Sun Yat-sen University is conveniently located close to Hong Kong and Macau, and has strongly pursued collaborations in those regions. Moreover, the university has forged long-term strategic partnerships with

institutions such as the Consortium Français de Formation d'Ingénieurs en Nucléaire Civil, Carnegie Mellon University and the Johns Hopkins School of Medicine. These collaborations have considerably extended the influence and enhanced the competitiveness of Sun Yat-sen University worldwide.

RESEARCH

Sun Yat-sen University enjoys excellent funding for scientific research, with annual government grants of over one billion RMB for longitudinal research for the past four years. Researchers have published thousands of high-quality articles in journals such as *Nature*, *Science*, *The Lancet* and *Cell* as well as their subsidiary journals. The university is ranked 12th in the Nature Index for China.

Achievements in academic research and consulting with government agencies have established the university's reputation both domestically and internationally. Globally influential research includes ecological studies where it was discovered that previous extinction rate estimates were overestimates — the actual rate of species extinction is roughly 40 per cent of previous estimates. This result was published in *Nature* in 2011 and overturned the former consensus in the scientific community regarding the species extinction rate. Another example of the university's groundbreaking research is the explanation of how the composition and structure of DNA in a cell affect spontaneous mutations. This research was published in *Science* in 2012 and provides important guidelines for understanding the genomic structure and evolution of eukaryotes. In 2014, it commissioned a

supercomputer with the fastest computational speed in the world, Tianhe-2, which will provide powerful support for cutting-edge research.

Sun Yat-sen University is strong in both traditional and newly emerging disciplines, including the classic basic sciences, information technology, materials science, health sciences, humanities, business administration and social sciences, and it has developed long-term, ambitious goals for all these fields. It has also drafted ambitious institutional development strategies for establishing a platform to conduct world-class research, including both large-scale recruitment and cultivation of internationally outstanding scientists and scholars.

RECRUITMENT OF TALENTED RESEARCHERS

The recruitment of talented teachers and researchers is crucial to the development strategy of Sun Yat-sen University. The university is very inclusive when recruiting outstanding researchers from both China and overseas. The university has commenced the second phase of its 100 Top Talents Program and hopes to attract exceptional researchers by offering excellent starting conditions, salaries and benefits and a free intellectual environment as well as transparent and efficient recruiting procedures.

Sun Yat-sen University is an ideal environment for both work and life, choosing it will make your dreams come true.



LEARN MORE

Visit: <http://rsc.sysu.edu.cn/Article/invitation/Professor/Index.html>

Phone: +86-20-84114884

e-mail: rscrcj@mail.sysu.edu.cn

Germany's top technical university — a global career destination

Discover TUM. The Entrepreneurial University.

The Technische Universität München (TUM) is one of Europe's leading universities of science and technology, engaged with society and open to the world. Consistently ranked by far number 1 among Germany's technical universities, TUM is taking action to become one of the world's most rewarding career destinations as well – for students, doctoral candidates, postdocs and faculty alike. As President, I want to extend my personal invitation: Find out more about Germany's "entrepreneurial university" in Munich, and consider the unique opportunities it offers.

Why Germany? Why not?

It's easy to say why Germany should be attractive to bright, talented people from all over the world. Our economy is strong and largely based on scientific discovery and technological innovation. Science is valued and broadly supported by our society. Students can get a world-class education here without taking on a crushing burden of debt. Researchers have access to funding sources that are well endowed and managed progressively, with a long-term perspective. The quality of life, by any measure, is excellent.

And yet we know there are barriers that can stand in the way. One of these, the perception that the German language is hard to learn, is understandable but out of date. We offer dozens of degree programs in English, more every year. This benefits not only international students but also young Germans competing in a global labor market where English is the *lingua franca*.

There is a more serious barrier, which mainly concerns those considering academic careers: the perception that the way German universities recruit and promote professors is not transparent, competitive, or fair. Here too, however, things are changing. We are leading the change.

Germany's first true tenure track, and more

TUM has introduced an end-to-end recruiting and career system that is consistent with best practices worldwide and remains – even as others begin to follow our lead – unique in Germany.

The difference can be seen particularly in terms of what TUM offers the world's most promising *young* researchers: real opportunities, based above all on performance, to achieve scientific independence and professional security. This system is flanked by policies and services that make it easier for international professors and their (often dual-career) families to make themselves at home in Munich – one of the world's best places to live and work, better still with a support network ready to welcome and assist newcomers.

The main thrust of this effort is the TUM Faculty Tenure Track System, launched in 2012 with a commitment to hire 100 new tenure track professors by 2020. With this TUM aims to force the transformation of the conventional German *appointment system*, a maze-like anachronism choked with obstacles, into an open, performance-oriented *career path*. This culture-changing enterprise is grounded in the recognition that the best young talents can only be recruited where they know they can advance within the system. Performance-based criteria – for the initial appointment of an *assistant professor* as well as for advancement to *associate professor* and *full professor* – are clearly defined and transparently implemented.

Our latest step in this direction offers the chance to win a joint appointment at TUM and one of the world's most prominent non-university research institutions, the Max Planck Society (MPG). A successful applicant to head a Max Planck or Minerva Research Group can simultaneously be appointed to a tenure track professorship at TUM. Thus together, MPG and TUM can offer a comprehensive package for promising young scientists – from first-class laboratory facilities to the right to confer doctorates and the option of further advancement as part of our faculty – that is (again) unique in Germany. MPG and TUM complement each other perfectly, offering both excellent working conditions and clear opportunities for career advancement right from the start.

TUM Faculty Tenure Track – together with initiatives such as senior-level professorships reserved for distinguished *women* in research, mechanisms to bring international scientists to Munich



Wolfgang A. Herrmann

PRESIDENT

Technische Universität München

as postdocs or as visiting professors, and a suite of family-friendly support programs – should make TUM an increasingly attractive destination for top talent worldwide. This in turn will be a key to our university's continued scientific excellence and its effectiveness as a servant of society.

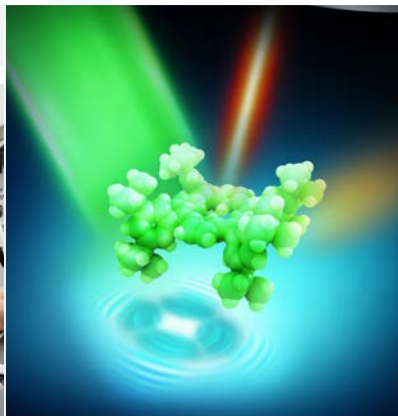
Make a difference in the world – your way

From students to senior researchers and industry partners, being part of TUM means having the chance and the means to make a difference in the world. Germany's technology-oriented universities have always had a core mission, dating back to the 1860s in our case, that is based on fundamental frontier research but strongly directed toward practical results. Basic insights spur innovations and new technologies enable discoveries, in an endlessly creative cycle.

Among the German technical universities, TUM commands the broadest portfolio of disciplines, covering natural and life sciences, engineering sciences, medicine, and economics. TUM fosters research and teaching at the highest level in all of these disciplines. We strongly promote interdisciplinary exploration, cultivate collaborations with industry and other partners, and actively prepare researchers for success in founding startup ventures.

In research, teaching, and partnerships as well as in public outreach and dialogue, today's TUM is oriented toward the major challenges facing society: energy, climate and environment, natural resources, health and nutrition, communication and information, mobility, and infrastructure. Addressing society's challenges requires interdisciplinary approaches that transcend boundaries of all kinds, now more than ever. That's the way we do it at today's TUM. If that works for you, think about joining us – take a closer look!

TUM. The Entrepreneurial University. Get to know us better at www.tum.de



THE UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA PURSUING EXCELLENCE IN SCIENCE

The University of Science and Technology of China (USTC) is one of the most important innovation centres in the country, and is always ranked among its best universities. It is particularly strong in fields such as quantum manipulation, nanotechnology, high-temperature superconductivity, speech processing, fire science and life sciences.

The USTC takes the lead in many major science projects, such as quantum satellite research and dark-matter detection. It is also an active contributor to significant international projects, such as the International Thermonuclear Experimental Reactor (ITER) and the European Organization for Nuclear Research (CERN).

In 2013, the USTC won more than 20 renowned awards in science and technology. For example, a team of USTC physicists led by Professor Xianhui Chen received the first prize in Chinese Natural Science for their contributions to the field of superconducting materials; for the previous three years, there had been no recipients of this prize.

Some of the latest research highlights are described below.

PHYSICS AND CHEMISTRY

High-energy physics at the particle colliders

A team led by Professor Zhengguo Zhao in the School of Physical Sciences made weighty contributions to the study of

diboson production, triple-gauge boson couplings and the discovery of Higgs particles via the ATLAS experiment at the Large Hadron Collider (LHC) of CERN. Zhao also greatly contributed to the observation of the Zc particles that were suggested to represent the charmed multi-quark states, using the Beijing Spectrometer (BESIII) at the Beijing Electron Positron Collider (BEPCII), and, for the first time, observed over 10 new decay modes of the charmonium states cJ and c. As a result of these outstanding achievements, Zhao was elected as an academician of the Chinese Academy of Sciences (CAS), which is the highest academic honour in the country.

Inorganic solid-state chemistry

Professor Yi Xie and her group at the Hefei National Laboratory for Physical Sciences at the Microscale (HFNL) pioneered research into the design and synthesis of inorganic functional solids with efforts to modulate their electron and phonon structures. Xie established the methodology known as the "synergetic use of binary characteristic structures" for the synthesis and assembly of inorganic functional materials, proposed a strategy for modulating the electron and phonon transport properties with phase transitions at the nanoscale, developed new high-efficiency thermoelectric materials systems, and discovered the relationship between the fine/electronic structures and the thermoelectric/optoelectronic properties of two-dimensional semiconductor crystals. As a female scientist, Xie is the youngest academician of the CAS among those elected in 2013.

Carbon aerogels sop up hydrocarbons

A team led by Professor Shuhong Yu at the HFNL is pursuing carbon aerogel



LEARN MORE

Visit: <http://en.ustc.edu.cn>
Phone: +86-(0)551-63607981



production from biomass. The team selected bacterial cellulose pellicles — a commonly used, inexpensive, nontoxic form of biomass consisting of a tangled network of cellulose nano fibres — as a precursor for the production of ultralight carbon nanofibre aerogels on a largescale. This biomass can easily be produced on an industrial scale through microbial fermentation.

QUANTUM INFORMATION AND QUANTUM TECHNOLOGY

The Synergetic Innovation Centre for Quantum Information and Quantum Physics (SIC-QIQP), head by Professor Jianwei Pan, was established and financially supported by the Chinese Ministry of Education. It focuses on bringing together teams of multidisciplinary researchers to form a dynamic national network for developing scalable quantum technologies.

Foiling quantum hackers

A research team led by Professor Qiang Zhang and Professor Tengyun Chen at the SIC-QIQP successfully demonstrated the measurement-device-independent

quantum key distribution by developing up-conversion single-photon detectors with high efficiency and low noise. The new quantum-encryption method provides the ultimate security against hackers in real-world cryptography applications, and greatly improves the security of quantum-encryption systems. This research was selected as one of the Highlights of the Year in *Physics* by the American Physical Society.

A milestone in satellite-based quantum communication

A collaborative team led by Professor Chengzhi Peng at the SIC-QIQP achieved comprehensive and direct verification of quantum communication between satellites and ground stations. This research lays the necessary technical foundations for a global quantum-communication network based on ground-satellite quantum communication by launching the quantum science experimental satellite of China.

Optical spectroscopy goes intramolecular

A team led by Professor Zhenchao Dong at the SIC-QIQP reported an optical spectroscopic-imaging approach that achieves subnanometre resolution and resolves the internal structure of single molecules. This development could lead to new techniques for probing and controlling nanoscale structure, dynamics, mechanics and chemistry. This research was listed among China's top 10 science news stories in 2013.

ENVIRONMENTAL AND EARTH SCIENCES

Penguins thrived in Antarctica during the Little Ice Age

New research led by Professor Liguang Sun in the School of Earth and Space Sciences showed that penguin populations in the Ross Sea of Antarctica spiked during the short cold period, called the Little Ice Age, which occurred between AD1500 and 1800. These results run contrary to previous studies that found increases in Antarctic penguin populations during warmer periods and decreases during

colder periods, suggesting that populations living at different latitudes in the Antarctic might respond differently to climate change.

Uncovering the mystery of subduction zone earthquakes

Based on analytical data from four of the highest magnitude subduction zone megathrust earthquakes, the conclusion was drawn that low-frequency radiation is closer to the trench at shallower depths and high-frequency radiation is farther from the trench at greater depths, in general. This scientific breakthrough was achieved by a team led by Professor Huajian Yao.

LIFE SCIENCES

New evidence for curing type 2 diabetes

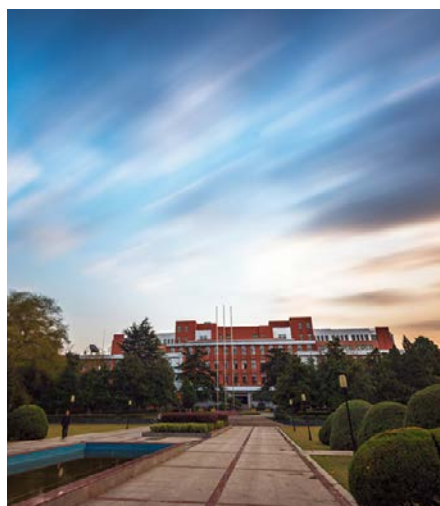
Research teams led by Professor Rongbin Zhou and Professor Zhigang Tian in the School of Life Sciences revealed a new mechanism through which omega-3 fatty acids inhibit inflammation and prevent type 2 diabetes. The research results were published in *Immunity* in June 2013 and highlighted in the same issue of the journal.

Identifying liver-resident natural-killer cells with immune memory

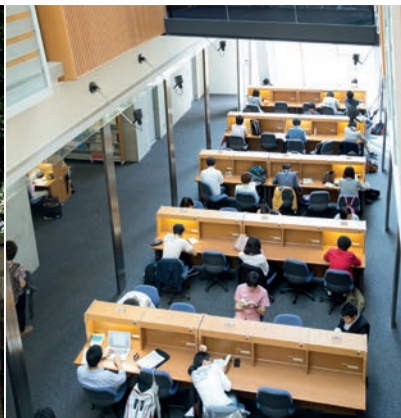
A team also led by Professor Zhigang Tian identified liver-resident natural-killer (NK) cells that possess unique immune memory characteristics absent from normal NK cells.

LincRNA-p21 as a novel key player in regulating the Warburg effect

A research team led by Professor Mian Wu and Professor Yide Mei, at HFNL and the School of Life Sciences, has revealed a novel mechanism whereby lincRNA-p21 regulates the Warburg effect under hypoxic conditions. They demonstrated, for the first time, that lincRNA-p21 is an important regulator of the Warburg effect, and also identify lincRNA-p21 as a valuable therapeutic target for cancer.



Guo Moruo Square. USTC was established by the Chinese Academy of Sciences (CAS) in 1958 in Beijing. The director of CAS, Mr. Guo Moruo was appointed the first president of USTC.



THE UNIVERSITY OF TOKYO RESEARCH WITH A BANG

Whether plumbing the depths of space to unravel questions of scientific and philosophical importance or contributing solutions to the most pressing problems that face Japanese society, researchers at the University of Tokyo strive to make their science count. The following three initiatives show what can be achieved through the university's unique commitment to interdisciplinary work and innovation.

The Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), bridges mathematics, astronomy, theoretical and experimental physics and has kept up a frenetic pace of achievement. Mathematician Yukinobu Toda's recent breakthrough proving the equivalence of two geometric theories of space marked an important step in the elucidation of string theory, and landed him a lecture at the International Congress of Mathematicians this year. Robert Quimby led a group of astronomers, physicists and a mathematician in correcting the scientific record—showing that claims of an unprecedentedly bright supernova by Harvard University researchers were actually the result of a gravitational lensing effect by a faint galaxy, which had not been previously detected.

Theoretical physicist Hiroshi Ooguri created a hot topic in mathematics with his discovery of the mysterious relationship between the geometry of a special four-dimensional space and a big discrete group with 244823040 elements—now known as 'Mathieu Moonshine'. In experimental physics, Alexander Kozlov joined a race to demonstrate the conversion between matter and anti-matter using

the Kamioka liquid scintillator antineutrino detector (KamLAND). The KamLAND-ZEN collaboration produced the world's most precise data to date. Hitoshi Murayama, Kavli IPMU's director, discovered a unified theory of the low-temperature limits of many systems ranging from magnets, superfluids, cold atoms, to neutron stars and cosmic defects, fifty years after various elements of these system were first theorized.

Kavli IPMU's research activities are set to expand beyond this extensive, crosscutting research base. The Subaru measurement of images and redshifts (SuMIRe) project, led by Kavli IPMU, has started to produce wide-field images of high-redshift galaxies using the Hyper Suprime-Cam on the Subaru Telescope, and a new robotic multi-fibre spectrograph, now under construction, will allow simultaneous observation of 2400 astronomical targets, making possible unprecedented surveys. Kavli IPMU also initiated a collaboration for a next-generation satellite, named LiteBIRD (which stands for 'Light satellite for the studies of B-mode polarization and inflation from cosmic background radiation detection'), which will deliver evidence for cosmic inflation and measure its characteristics. Murayama also plans to add statistics as another focal point, which "will create new synergies between astronomy and mathematics."

The institute has also been spreading creative ideas about how to organize and support science. To a rigid Japanese system, Murayama has introduced split-appointments, merit-based salaries, and lengthy, negotiated recruitment procedures, all of which made it possible to assemble top talent from around the globe. "This is very different from what is traditionally done in Japan," says Murayama. For an institute that gets nearly 1000 applications every year, more than 80% of which are from abroad, it is also a tremendous effort. Murayama



THE UNIVERSITY OF TOKYO

LEARN MORE

Visit: www.u-tokyo.ac.jp/en
www.facebook.com/UTokyoNews.en
www.twitter.com/UTokyo_News_en
www.youtube.com/user/UTokyoPR
 Phone: +81-3-3811-3393



intends to keep stretching the boundaries of Japanese academic research with a new international graduate program.

A radically new organizational system, called the Todai Institute for Advanced Study (TODIAS) introduced in 2011, helped Kavli IPMU achieve the flexible system that enabled its growth. The TODIAS system then opened the path for another ambitious, interdisciplinary initiative, one targeting a topic of grave concern.

The Integrated Research System for Sustainability Science (IR3S), the second TODIAS institute, “was founded with the vision of providing cutting-edge research for building a sustainable society through linking global, social and human dimensions,” says director Kazuhiko Takeuchi. Takeuchi is convinced “that the problems our society currently faces, such as climate change, biodiversity loss, and poverty, are complicated, rapidly changing and difficult to solve through the ‘single mindsets’ of existing academic disciplines.”

IR3S breaks down these walls, and its researchers, who represent fields ranging from ecology and environmental sciences, to economics, engineering and public health, are involved in highly interdisciplinary research. Their aim is to become a driver of social innovation and catalyze reform within other scientific fields. Ongoing studies, which have taken IR3S researchers across Asia and Africa, are as diverse as adaptation and resilience to climate and ecosystems change, health and global change, the impact of bioenergy, agricultural sustainability, and green economy.

“In all these projects we routinely work with local partners and adopt a problem-oriented and interdisciplinary approach that considers the information needs of end-users,” says Takeuchi. During a recent visit to Ghana, for example, IR3S researchers communicated findings that will help local communities deal with increased flood and drought hazards. The published results of these projects have been used to formulate policy advice to local and international stakeholders.

This global vision, inherent in an institute where 70 per cent of the workforce is

international (compared to fewer than 10 per cent in most Japanese universities), is further bolstered by high-powered advisors, including Rajendra K. Pachauri, chairman of the Nobel Peace Prize winning Intergovernmental Panel on Climate Change.

IR3S recently entered a strategic collaboration with the Science Council of Japan, to host one of five globally distributed secretariats of Future Earth, an international initiative that will coordinate interdisciplinary research efforts globally. Involvement with Future Earth will allow IR3S “to set the research agenda, fundraise, and build capacity for sustainability research globally in the next decade,” says Kensuke Fukushi, Interim Director of the Tokyo hub of Future Earth Secretariat. IR3S is also establishing an Asia-wide university network on climate change mitigation, called the 3E Nexus Initiative whose aim is to assist the implementation of the Japanese Government’s Joint Crediting Mechanism (JCM) in countries of Asia-Pacific. Involvement in other initiatives, including the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) and the International Society for Sustainability Science (ISSS), further highlight the leading role that IR3S plays within the international sustainability science community.

The university’s innovative initiatives are also tackling urgent problems closer to home. Facing a declining birthrate and increasing longevity, in 2013, the government awarded the university funds to form a Center of Innovation (COI) entitled “Sustainable life care, ageless society” whose mission, says director Tomihisa Ikeura, is to “realize a sustainable society where people can live longer in good health, thus alleviating the burden of the younger generation and allowing the elderly and younger populations to happily coexist and prosper.”

The center’s efforts will fall on three intersecting axes: minimally-invasive diagnosis and treatment; high-performance and low-cost drug discovery; and the health care and sick care data platform.

The University of Tokyo provides a fertile ground in which to realize such a vision. The center will take advantage of the university’s

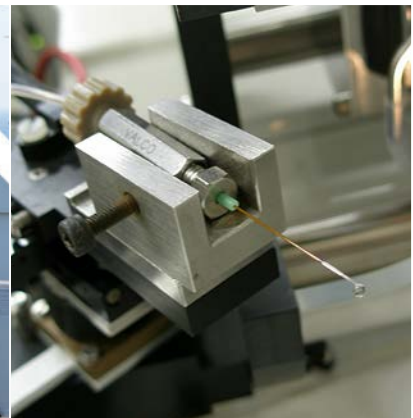
outstanding seeds of innovative science and technology, the scientific resources at the medical, engineering, and pharmaceutical science schools, the clinical resources at its hospital, and its ability to provide stable funding that reduces the risk for the many member companies who have joint research projects. Flexible hiring policies, including laxer regulations on dual appointments, has helped the center land elite staff.

Ikeura, who was a managing executive officer at Mitsubishi Chemical Holdings, is well aware of the need for supportive regulatory framework, and the center will also collaborate with the Pharmaceuticals and Medical Devices Agency and the National Institutes of Health Sciences (NIHS) to streamline the regulatory path.

To encourage cooperation among the sectors, the center launched the “Healthy Long Life Loop Society” which will invite both COI member and non-member companies spanning from electronics, mechanical equipment, information and technology, chemical and pharmaceutical technologies to food, transportation, nursing care, fitness, and life insurance, in order to exchange information related to health care, medical device regulation, and other relevant matters. “This will create an environment where new businesses, new inter-business alliances, joint businesses, overseas research and business expansion opportunities are spontaneously fostered,” says Ikeura.

Ikeura hopes that eventually the center will, as an independent and sustainable organization within the University of Tokyo, realize a fundamentally novel medical system, one “that dramatically reduces costs of medical and nursing care, creates new markets and globally competitive new medical industries, and establishes a sense of security through the bonding of people across generations.” When it achieves this, it will no doubt become a model for other countries grappling with similar problems.

While these initiatives are all anchored in research groups situated in Japan, like the university’s hundreds of other cutting-edge scientific projects, their influence reaches far beyond Japan’s borders.



TOKYO UNIVERSITY OF AGRICULTURE AND TECHNOLOGY GLOBAL INNOVATION RESEARCH ORGANIZATION FOCUSES ON THE AREAS OF FOOD, ENERGY AND LIFE SCIENCE

Tokyo University of Agriculture and Technology (TUAT) has a long history of improving human society through advances in agricultural science and engineering. Ever since its beginnings in the 1870s as two government departments that merged in 1949 to form a university, TUAT has placed a strong emphasis on improving the lives of people in Japan and beyond.

That emphasis received an international twist this year with the launch of the Global Innovation Research Organization, which will bring together TUAT researchers and leading international professors to tackle global problems in the areas of food and energy. Measures include increasing crop yields to combat global food shortages, monitoring air pollution in East Asia and safeguarding the world's energy supply.

The creation of the Global Innovation Research Organization was the brainchild of the University's president, Professor Tadashi Matsunaga, and it was supported by priority allocation from the national universities' budget. It promises to transform the way research is undertaken at TUAT. The initiative will increase connections with international partners and provide novel management training opportunities for young researchers in the areas of food and energy.

Led by the vice president of TUAT, Professor Chisato Miyaura, the Global Innovation Research Organization will focus on three critical priority areas: food, energy and life science. These areas reflect TUAT's focus on addressing the most immediate problems facing humanity. Each area will have three teams, which will be led by a TUAT researcher teamed with a distinguished international professor. The international professors will be invited to participate in the organization as recognition of their high standing in the global scientific community, and they will visit TUAT for one to three months to focus solely on research.

FEEDING THE WORLD

Food shortage is a devastating global issue, affecting 842 million people, 66% of whom are in the Asia-Pacific region. Hence, food is one of the research priority areas of the Global Innovation Research Organization. These research programmes have the potential to positively impact our food supply, health and the environment.

“In these collaborations with distinguished professors from around the world, the Global Innovation Research Organization will foster such abilities in young researchers.”

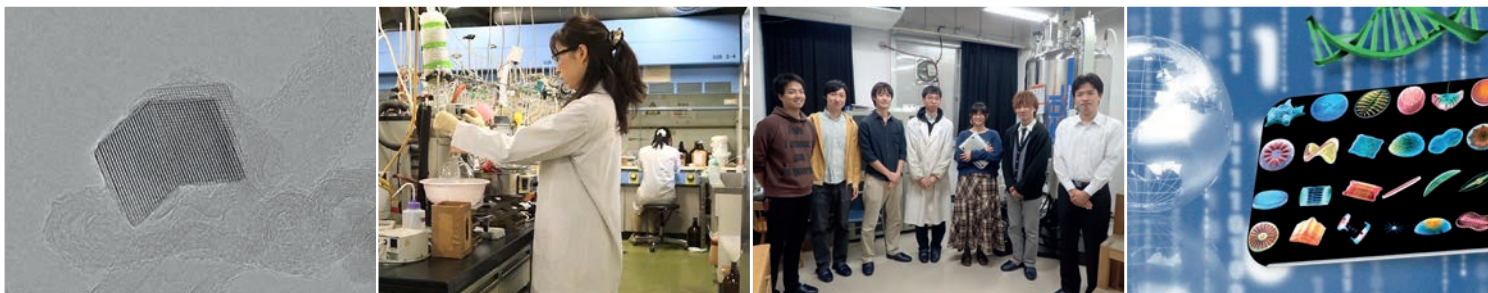


Tokyo University of Agriculture and Technology

LEARN MORE

Visit: www.tuat.ac.jp/en/
E-mail: kenkyu1@cc.tuat.ac.jp

The first team in the food group is led by Professor Tadashi Hirasawa, who, in collaboration with a distinguished professor from the USA, will focus on using genome information to increase the production yield of a variety of crops. Hirasawa's team will identify which genes in plants facilitate photosynthesis



and then use this knowledge to propose a selection process designed to enhance crop yield. Professor Nobuhiro Takahashi leads the second team, which will focus on studying the turnover and physiological function of RNA in animal cells. "The insights gained from this work have the potential to provide novel targets for the early diagnosis and treatment of severe neurodegenerative diseases," he says. Meanwhile, Professor Shiro Hatakeyama's team, which consists of a network of collaborators in Japan, Korea, Taiwan and Hong Kong, will study transboundary air pollution in East Asia. The collaboration with the distinguished international professors will assist the research teams in understanding how air pollutants are distributed and transformed across the East Asia region and clarify how specific emission sources contribute to air pollution.

ENERGIZING THE GLOBE

Energy production and storage are becoming increasingly important topics, with world energy consumption predicted to increase by 56% between 2010 and 2040. Hence, the second priority area of the Global Innovation Research Organization is energy. Professor Hiroyuki Ohno leads the first team, which will use ionic liquids to develop novel energy conversion technologies in collaboration with a distinguished professor from Italy. "We will focus specifically on improving the performance of batteries, fuel cells, biorefinery processes and novel refrigeration processes," he says. Taking another approach, Professor

Yoshinao Kumagai leads the second team, which will seek to develop materials for low-loss and high-power electrical devices. In collaboration with a distinguished professor from Sweden, Kumagai's team will concentrate on developing high-quality, crystalline group-III (for example, metals such as aluminium, gallium and indium) nitride semiconductors. Advanced supercapacitors is the topic of choice for Professor Katsuhiko Naoi and the third team in the energy group. These researchers, which include a distinguished professor from France, have proposed a new supercapacitor system, dubbed the 'nano-hybrid capacitor', so called because it is based on a combination of carbon nanofibres and nanocrystalline lithium titanate. "We believe that the nano-hybrid capacitor could help meet the energy and power demands for a variety of applications, including microelectronic devices and electrical vehicles; it has the potential to be a breakthrough improvement," Professor Naoi says.

INVESTIGATING NATURAL SYSTEMS

The life science priority area brings together three very different but thematically linked investigations. The first team is led by Associate Professor Masaki Inada who has teamed up with collaborators from Spain, the UK and Germany. This team will utilize the respective expertise of Inada and the distinguished professors to develop biomedical applications for treating diseases such as cancer based on complexes of

collagen molecules. The second team, led by Professor Kazuhiro Chiba with a distinguished professor from Finland will focus on developing nanoparticle drug delivery systems, combining the preparation of a range of nanoparticles and biologically active complexes to develop nano-sized drugs capable of being absorbed through the skin. The third team, led by Associate Professor Tsuyoshi Tanaka with input from distinguished professors from France and the USA will establish a new research group of excellence for 'marine-omics'. This field combines data from genomics, environmental microbiology, geochemistry and molecular biology to provide rich insights into marine microalgae and how they may be used to address current issues in food, energy and health research.

TRAINING THE NEXT GENERATION OF RESEARCHERS

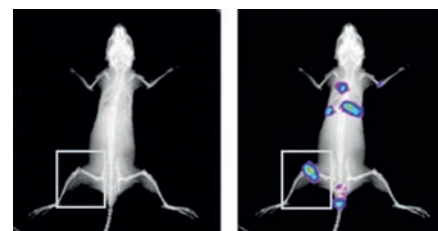
In addition to tackling challenging scientific problems, the Global Innovation Research Organization aims to help scientists, especially early career researchers, become literate in management principles. "It is unusual to find young researchers in Japan who have an appreciation of both scientific techniques and management skills," remarks Professor Miyaura, the head of the organization. "In these collaborations with distinguished professors from around the world, the Global Innovation Research Organization will foster such abilities in young researchers." In this way, TUAT aims to promote the globalization of research in its focus areas, especially among young researchers.



Genetic analyses in plants facilitate photosynthesis to enhance crop yield.



Growth reactors of new single crystals for low-loss and high-power electrical devices



Bio-imaging analyses for medical application of diseases such as cancer.

Defining the Future of the
Public Research University

UC San Diego

Transcending Health-Care Boundaries

Human health care is advanced by crossing boundaries: cell walls, technological limits and political borders. Traversing these boundaries is critical to quality care and medicine.

At UC San Diego, our award-winning faculty cross barriers to revolutionize how, where and when medical care is delivered. Improving health on all scales—from the nano to the global—demands collaboration across disciplines. From bench to bedside, our work leads to innovative treatments that touch the lives of people worldwide.

Jacobs Medical Center, scheduled to open in 2016, will house 245 patient beds and include three specialized hospitals that represent the future of patient care in the region. The facility will be the anchor of the La Jolla medical campus of UC San Diego Health System.



Steffanie Strathdee, Ph.D.

Grassroots Global Health

Fighting HIV and sexually transmitted infections means hitting the streets.

Steffanie Strathdee, Ph.D., mobilizes binational collaborations and students to reach out to drug users and sex workers on the Mexico-U.S. border. Her work with these underserved and marginalized populations, local police, federal authorities and nongovernmental organizations provides critical information for public policy.



Adah Almutairi, Ph.D.

Exploring the Frontier of Drug Delivery

Adah Almutairi, Ph.D., directs the Center for Excellence in Nanomedicine and Engineering. Researchers in the center are creating nanoparticles that promise new levels of precision in treating disease. They will be able to deliver drugs under highly specific conditions—when triggered by light or when they encounter a disease-related condition.



Lawrence Goldstein, Ph.D.

Tackling Alzheimer's With Stem Cells

Lawrence Goldstein, Ph.D., director of the new Sanford Stem Cell Clinical Center and scientific director of the Sanford Consortium for Regenerative Medicine, has created, for the first time, stem cell-derived models of sporadic and hereditary Alzheimer's disease using induced pluripotent stem cells from patients with AD. These functional neurons in a dish promise to be an unprecedented tool for developing and testing drugs to treat patients.

#1

Among San Diego's adult hospitals:
UC San Diego Health System
2012-13 U.S. News & World Report
"America's Best Hospitals" issue

UC San Diego

Learn more at ucsd.edu.



:insideview

profile feature



Professor Yonghua Song, Executive Vice-President, Zhejiang University, Hangzhou, China

Located in the historical and picturesque city of Hangzhou, Zhejiang University is perfectly positioned to be an innovative, comprehensive research university with world-class standing. Throughout its 117-year history, Zhejiang University has maintained a commitment towards cultivating talent, with the primary goals of pursuing excellence, advancing science and technology, developing society and promoting culture. This spirit is manifested in the university motto “Seeking the Truth and Pioneering New Trails”. Zhejiang University is a comprehensive research university with national and international reach. Research at Zhejiang University spans 12 academic disciplines, including philosophy, economics, law, education, literature, history, art, science, engineering, agriculture, medicine and management. In 2013, the research fund at Zhejiang University totalled 3 billion RMB. Moreover, 116 research projects have each secured a grant of more than 10 million RMB.

Q: What are Zhejiang University's goals?

Zhejiang University is the first university in China solely founded by Chinese nationals. From its inception, Zhejiang University's pedagogical approach was characterized by practical application and societal impact. In the past 20 years, Zhejiang University has aimed to become a world-class university, taking the lead among universities in China to explore and exercise a training system for top-level innovative talents in the establishment of a research university.

We select only the best students from across the country for freshman class. We are also the most selective graduate school in China. We want every Zhejiang University professor to be able to create innovative knowledge and train students to create such knowledge, laying the groundwork for every graduate to demonstrate leadership qualities and make a positive impact on society. This is the most fundamental mission of Zhejiang University.

Q: What are Zhejiang University's key strengths?

Zhejiang University covers the most diverse areas of study of all universities in China and is ranked number 159 in academic influence by the Essential Science Indicator (ESI) ranking of September 2014. With 17 areas of study ranked in the top 1% in the ESI, Zhejiang University is the most rapidly improving university in China. Upside potential of Zhejiang University is not only because of the China's rapid economic development, but also because the Zhejiang province is a highly innovative and economically dynamic region with an annual GDP of 3.75 trillion RMB (612 billion USD).

Hangzhou, the capital of Zhejiang Province, is a cultural and scenic city, described by the Italian explorer Marco Polo as “the most beautiful and most attractive town”. Alibaba, a world-renowned Chinese enterprise is headquartered in Hangzhou. These geographic advantages also contribute to the growing potential of Zhejiang University.

Q: What qualities does Zhejiang University look for when hiring researchers?

Each year, well-known scholars from academic institutions around the world choose Zhejiang University as a home where they can express their strong academic creativity. We hope that the most competent young scholars from a range of research fields will join Zhejiang University, leading their research teams to produce internationally competitive academic achievements, and using their charm and ability to influence and train young students.

Q: What does Zhejiang University do to attract top academic talent?

In recent years, Zhejiang University has been committed to building an academic environment designed to attract talent. These commitments have involved developing our policy and funding programmes. We have been able to provide a high standard of living and international

academic research conditions for outstanding academics.

The University has set aside 1.8 billion RMB to invest in its high-level faculty team over the next 4 years, which is very rare among Chinese universities. Zhejiang University has implemented the tenure track system to provide young researchers with quality academic conditions and wages. We are hoping that this will significantly enhance the development of Zhejiang University.

“Hangzhou, the capital of Zhejiang Province, is a cultural and scenic city, described by the Italian explorer Marco Polo as “the most beautiful and most attractive town”

Q: What do you consider to be the main challenges for Zhejiang University to become a world-class university?

I studied and worked in the UK for 18 years, prior to which I had gained 20 years of research experience in China. These experiences were important for my personal growth as well as my academic and administrative capacity; they also allow me to view different academic and educational institutions objectively. I believe the key to building a world-class university is to gather a team of international faculty.

But by “international” I do not mean that the focus has to be on the actual nationalities of the faculty members; instead it refers to an open mind and the ability to communicate with the world.

www.zju.edu.cn/english

Zhejiang University · 866 Yuhangtang Road · Hangzhou · Zhejiang Province · 310058 · P. R. China

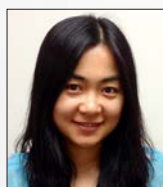
Zhejiang University

SEEKING THE TRUTH AND PIONEERING NEW TRAILS

As a national university, Zhejiang University is building itself into an innovative and comprehensive research institution of world class standing. Zhejiang University adheres to the spirit of seeking the truth and pioneering new trails, dedicating itself to creating and disseminating knowledge, inheriting and advancing civilization, serving and leading society, and promoting national prosperity, social development and human progress.



Why did you choose Zhejiang University?



"Zhejiang University acts as a bridge to the exciting and fast-growing Chinese market. It allows researchers to harvest real-life problems and provides abundant

opportunities for applying research outcomes to real systems. Besides providing opportunities to work with the best students in China, Zhejiang also has an excellent research environment that nurtures world-class research."

Professor Wen Yuan Xu, College of Electrical Engineering



"For a young researcher, China is a fascinating country to begin an independent career. The creativity and pace of modernization will impress everyone who visits,

and having a position here gives you the opportunity to focus on what really matters: research and the education of a new generation of scientists. Zhejiang University is one of the top universities in China and provides starting faculty with excellent working conditions.

The Department of Chemistry supports its young professors wherever it can, gives them much freedom and welcomes interdisciplinary collaboration. I am happy to have brilliant and helpful colleagues as well as smart, hard-working students."

Research Professor Simon Duttwyler, Department of Chemistry



"As a former PhD student at Zhejiang University, I have always valued its strong support network, so after doing two postdoc programmes

here, when the opportunity came for me to stay and teach, I took it without hesitation. I am proud to be on the faculty of the College of Agriculture and Biotechnology and I feel like I can never repay my alma mater for transforming me from a young boy into the quality teacher that I am today."

Lecturer Imran Haider Shamsi (Ying Lan), Department of Agronomy



"I decided to return to China in 2009 after working abroad for 15 years as I believe that China is the best place to find sustainable devel-

opment solutions to address our planet's limited resources. Indeed, the Yangtze River Delta, where Zhejiang University is located, is an area that has prospered continuously for thousands of years without major environmental decay. In particular, I enjoy the freedom to pursue my area of interest as a researcher and the open-minded atmosphere at Zhejiang University."

Professor Xiaogang Peng, Department of Chemistry



"Although I officially joined Zhejiang University in 2011, I have been collaborating with researchers at Zhejiang University since 2002. I have found it to be a place

with great academic freedom, and I believe that's the reason why Zhejiang University is one of the best universities in China. In addition, I am grateful to be able to work in the beautiful city of Hangzhou, where I grew up. Zhejiang University is also strong in engineering, which provides a fantastic platform for physical science research and interdisciplinary collaboration — activities that are of particular importance to me."

Professor Fuchun Zhang, Department of Physics



"Zhejiang University's location in the economically dynamic region of the Yangtze River Delta provided much opportunity for translational research and a large patient base for biomedical research."

Professor Lanjuan Li, Director of the State Key Laboratory for Infectious Diseases Diagnosis and Treatment



浙江大学
ZHEJIANG UNIVERSITY